The Practicality and Validity of using Outcomes to Indicate the Quality of Stroke Care

Nicolas U. Weir
MBChB (Hons.) MRCP MSc

MD Thesis

The University of Edinburgh

2003

Contents

Declaration

I, Nicolas Weir, hereby certify that:

a. this thesis has been composed entirely by myself

b. that the work contained herein is my own work*, excepting those areas where the help of others is acknowledged

c. that I undertook the work contributing to this thesis whilst employed in the Department of Clinical Neurosciences at the University of Edinburgh, and

d. that I have not submitted this thesis in candidature for any other degree, postgraduate diploma or professional qualification.

Date 4th January 2004

* The thesis is based upon a collaborative study, the Stroke Outcomes Project, which involved various clinicians, statistical, computing and administrative staff. My specific contribution to the study is outlined in Appendix 12.

Acknowledgements

This thesis would not have been possible but for the help of many people. In particular, I would like to thank my supervisor, Dr. Martin Dennis, for conceiving and designing the Stroke Outcomes Project and for his wise counsel, encouragement and patience. I am also indebted to the members of the Scottish Stroke Outcomes Group (Dr Martin Dennis; Dr. Peter Langhorne; Dr. Jim McCallion; Dr. Jacqueline Taylor; and Professor Ian Campbell) who facilitated the collection of data at the five study hospitals, and to Steven Kendrick and the staff at the Information and Statistics Department of the Common Services Agency of the NHS in Scotland who provided the routine hospital discharge and case fatality data. I am also grateful to Alison Gunkel, our tireless and meticulous audit assistant; to Mike McDowall, who constructed and ran the study databases and provided invaluable computing support; to Pam McLaren and Marion Livingstone for day to day administration; to Dr Carl Counsell who acted as my partner in the prospective inter-rater study and to all who contributed to the prospective stroke register at Hospital D (the Lothian Stroke Register); to Stephanie Lewis and Dave Signorini for statistical advice; and to the Wellcome Trust who employed me during my period in research. Finally, I am deeply indebted and grateful to my wife, Stella, daughter, Katie, and son Alex (who waited and did not know) who supported me through the long and difficult process of writing this thesis. They have shown patience and understanding far beyond reasonable expectation. I dedicate this thesis to them.

Abstract

Routinely collected outcomes (case fatality by 30 days and discharge home by 56 days) have been used to indicate the quality of hospital stroke services in Scotland since 1994. However, the validity of these data is in doubt. In particular, it is difficult to know whether differences in outcome are due to differences in the quality of care or to inadequate adjustment for casemix, biased measurement or the play of chance. Also, because the majority of patients survive their stroke, the relevance and sensitivity of the indicators is limited by the failure to report functional outcome.

It would be useful, therefore, to investigate whether a substantially improved system (one which adjusts comparisons of outcome for important differences in casemix and which measures functional outcome at a defined time after admission) would be routinely feasible and provide valid measurements of the quality of stroke care.

We attempted to address these questions in the context of a study of 2724 patients with an acute stroke admitted to five Scottish hospitals between 1995 and 1997. We identified patients using routine hospital discharge information and then identified cases of acute stroke and data describing casemix and the process of care from the medical record. We collected case fatality by linkage to death certificate data and functional outcome and institutionalisation by postal survey at six months. We adjusted comparisons of outcome for casemix using a set of simple, externally validated logistic regression models.

The study shows that a considerably improved routine system for measuring outcome after stroke is a realistic possibility. Specifically, it suggests that the proposed system for identifying hospital cohorts and collecting casemix data would be reasonably accurate and that the proposed system for measuring functional outcome, although compromised by non-response, would not be seriously biased. Nonetheless, whether these improvements would result in valid measurements of the quality of stroke care remains uncertain. At best, it appears that a system reporting case fatality and death or dependency at six months might be sensitive to moderately large differences in the quality of care. However, there may be alternative explanations for this finding and the system would certainly fail to identify opportunities to improve care at the majority of hospitals. The collection of data describing simple but important aspects of the process of care in addition to outcome might be preferable and should be investigated.

Introduction

" People measure what is measurable and collect the results like jackdaws, regardless of value or usefulness"

Maxwell, 1992

Until not so very long ago, questions regarding the quality of clinical services were rarely raised in the UK. By and large, physicians were presumed to work to the best of their abilities to deliver high quality medical care, trust being placed in their training and professionalism to ensure that this was the case. Recently, however, the medical and social climate has changed and, for better or for worse, the trust of patients, public and government in the medical profession has been eroded. As a result, and combined with recent reforms of the NHS, physicians in the UK now find themselves subject to considerable pressures to demonstrate the quality of their work and, where necessary, show that they are making efforts to bring about improvements. The medical profession has itself contributed to this state of affairs, and, armed with its evidence-based methods, systematic reviews and clinical guidelines, it too has become interested in checking the quality of its own practice and in identifying opportunities to enhance and expand services.

Unfortunately, the simple convergence of opinion that the quality of care should be measured has not detracted from the fundamental complexities of the task. Whether the aim is to establish the quality of the environment of care (structure) or of the actions of clinicians and other staff (process), or to measure the net result of care (outcome) methodological pitfalls abound, especially when the system of measuring

the quality of care is to be applied routinely. The tendency to compromise measurement – for reasons of cost - by relying on sub-optimal data collection systems that already happen to be in place is a particular source of difficulty. Problems also arise from the purpose and context of the system of measurement. Systems which aim to ensure accountability can seem remote and carry notions of public scrutiny and punishment and so may not foster much in the way of quality improvement, whilst internal systems which aim to improve quality of care, although 'user friendly', run the risk of being quietly ignored. Thus, there has been and remains much debate as to the correct approach to measuring the quality of care and uncertainty as to the value of the mechanisms that are in place.

The Stroke Outcomes Project (SOP), the study upon which this thesis is based, was conceived in response to the establishment of an external system for measuring the quality of hospital stroke services in Scotland in 1994. Drawing on its legacy of (relatively) high quality routine hospital discharge information and its ability to link these data with centrally held death certification data, the Clinical Resource and Audit Group (CRAG) of the NHS in Scotland determined that these data should be used to measure and compare the outcomes of patients admitted to Scottish hospitals with an acute stroke. Specifically, the system compared 30 day case fatality and the proportion discharged home within 56 days of admission from there, adjusted for age, sex and socio-economic status, and its findings were made public. Despite their relative sophistication, the CRAG Stroke Outcome Indicators were received with scepticism by the medical profession. In particular, reservations remained about the accuracy with which the routine hospital discharge data were able to identify cases of

stroke, the failure to adjust the comparisons of outcome for clinically important differences in casemix (i.e. the failure to compare like with like), the failure to measure functional status (the outcome of most relevance to survivors) and the continuing impact of the play of chance. As a result, the system generated much debate about the pros and cons of quality measurement but little actual improvement in the quality of stroke care.

As a response to the publication of the CRAG Stroke Outcome Indicators, the purpose of the SOP was two-fold. First, the SOP aimed to determine whether the key shortcomings of the CRAG data could be addressed and hence whether a considerably improved system for measuring outcome after stroke could be made to work in the real world; and second, the SOP aimed to explore whether such an improved system might then act as a valid method for identifying hospitals with different standards of stroke care. The specific improvements to be tested were the routine collection of simple but important baseline data (to more adequately adjust the comparisons of outcome for differences in casemix) and the routine collection of functional status in survivors at a defined time after admission.

This thesis describes the SOP and attempts to place its findings into context. Chapter One describes the underlying concepts and approaches to measuring the quality of care in general, its historical background, the forces that drive the 'quality agenda' and its current focus on outcomes, and the considerable methodological hurdles faced by anyone wishing to use comparative outcomes data to indicate the quality of care. Chapter Two concentrates on the factors that have influenced the current drive

to measure the quality of care for patients with stroke and (non-systematically) reviews the literature describing the validity of using casemix-adjusted, observational outcomes data for this purpose. Chapter Three describes the methods of the SOP in detail.

The next three chapters address specific aspects of routine outcome measurement. Chapter Four explores the accuracy with which routine hospital discharge data are able to identify cases of acute stroke and estimates the impact of any inaccuracy on the comparison of outcomes; Chapter Five investigates the reliability and validity with which key baseline characteristics (used to adjust for casemix) can be collected; and Chapter Six tests simple strategies to improve response to follow up and estimates the impact of any non-response bias on comparisons of functional outcome.

The last three chapters relate to the SOP proper. Chapter Seven describes our success in identifying cases of acute stroke and in collecting casemix and outcomes data, and then compares the outcomes of the study hospitals before and after adjusting for casemix. Chapter Eight describes the findings of our survey of structure and process of care and relates them to the outcomes, adjusted for casemix. The final chapter, Chapter Nine, summarises the main findings of the SOP, suggests possible improvements to our proposed system and explores alternative approaches to the measurement of the quality of stroke care.

Glossary

A&E	Accident and Emergency Department
ADL	Activities of Daily Living
CI	Confidence Interval
CT	Computed Tomography
CVD	Cerebrovascular Disease
DepCat	Deprivation Category
GCS	Glasgow Coma Scale
GP	General Practitioner
GRO	General Register Office
ICD	International Classification of Diseases
ISD	Information & Statistics Department (of the Common Services Agency of the NHS in Scotland)
MRC	Medical Research Council
MDT	Multi-Disciplinary Team
OR	Odds Ratio
RCPSAP	Royal College of Physicians Stroke Audit Package
RR	Relative Risk
SAH	Subarachnoid Haemorrhage
SMR1	Scottish Morbidity Record (Type 1)
SRU	Stroke Rehabilitation Unit
TIA	Transient Ischaemic Attack

Chapter One: Quality of care and its measurement using outcomes

1.1 An approach to measuring the quality of health care

Defined by the Oxford English Dictionary, the term 'quality' refers to the degree of excellence of a thing. Given its many facets and complexities, modern health care may be considered to be excellent (or otherwise) in a number of different ways. The Institute of Medicine distilled the concerns of most observers into its single definition of the quality of care, namely 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge' (Blumenthal 1996a). This definition is useful in that it emphases the need to consider different perspectives, the need for quantification, the probabilistic relationship between care and its outcome and the importance of scientific knowledge (Miles *et al.* 1995). However, it remains a theoretical construct and, in order to make it operational, one must return to a multidimensional view. Maxwell suggested that the quality of care could be judged using a framework of six dimensions (Maxwell 1984):

Effectiveness. Is the best technical care being provided as judged by those with the requisite knowledge and experience?

Acceptability. Is the health care provided in a socially acceptable manner taking into account the views of patients and interested third parties?

Efficiency. Is the provision of care value for money?

Access. Are there barriers between patients and services, such as distance, ability to pay, waiting lists, waiting times or poor supply?

Equity. Are patients treated fairly in relation to others?

Relevance. Is the overall provision of services in line with the health care needs and desires of the population as a whole?

The framework can be used to guide the measurement of quality at all levels of care from the individual clinician through to the national level (Maxwell 1992). Different observers tend to be specifically interested in different dimensions: health professionals tend to be most interested in the effectiveness of care; funding bodies in efficiency; patients and families in acceptability and accessibility; and government in equity and relevance.

Performance within each dimension can be measured according to Donabedian's triad of structure, process and outcome (Donabedian 1988). Structure refers to the setting of care and encompasses its physical and human resources (e.g. equipment and services, the qualifications and numbers of staff) and its methods of organisation (e.g. of its clinical teams, volume of patients, training, quality control, funding, etc.). Process refers to what is actually done during the patients journey through the health care system and encompasses assessment, investigation, diagnosis and the provision of treatment. Outcome refers to the change in the health status of the patient that results. A vital point regarding any 'structure, process, outcome' triad is that its individual elements should be inter-linked. For questions of effectiveness and relevance, the evidence of these linkages should come from the medical sciences or,

failing that, from rigorously derived consensus statements of best practice. Thus, a process based measurement of the effectiveness of care is only valid if it is known (or strongly believed) that variation in the measured process leads to variation in outcome; similarly, an outcome based measurement of the effectiveness of care is only valid if it is known (or strongly believed) that variation in the measured outcome can be attributed to variation in the process of care (Brook *et al.* 1996; Donabedian 1988).

The relative merits of the structure, process and outcome triad are often debated. Structure is generally regarded as the least important, necessary but not sufficient to ensure the provision of high quality care, and as such viewed as a blunt tool more important to system design than the routine measurement of quality (Donabedian 1988). Evidence that associates potentially measurable structural elements such as numbers of nursing staff, organisational culture and patient volumes with health care outcomes has lead some to question this assumption (Knaus *et al.* 1986; McKee *et al.* 1997; Rudd *et al.* 2001a; Stroke Unit Trialists' Collaboration. 1997a). Nonetheless, the emphasis has remained on measurements of process and outcome, with each falling in and out of favour. In truth, the two approaches have different measurement properties and the decision as to which is the better to use depends very much on the nature of the care that is under investigation and the level at which the measurement is being made (Donabedian 1988; Mant 2001; Mant and Hicks 1996). Arguably, a more rounded view would result in any situation if both were measured (Donabedian 1988). However, for practical reasons this is not always possible. On the whole, where there is sound evidence that well-defined treatments are effective and

generalisable, it makes most sense to measure quality simply by measuring the process of care directly (Long 1997a; Mant and Hicks 1996). Here, rigorously derived evidence and consensus based clinical guidelines can be used to derive credible review criteria (Baker and Fraser 1995a; Grimshaw and Russell 1993). This is not to say that the approach is easy or unbiased (Gompertz *et al.* 1994a; Hulka and Romm 1979; Mannion and Davies 2002; Ryan and Dodd 1993), but it is the most direct, and hence it is the most sensitive and specific to variations in quality of care and its findings are the most simple to interpret (Crombie and Davies 1998; Mant and Hicks 1996; Palmer 1997). In many other situations where knowledge about the processes of care and/or their generalisability is less secure it may be more sensible to monitor patient outcomes (Long 1997a). This approach is, however, considerably less direct. Basic observational epidemiology reminds us that many factors other than the quality of the process of care, namely measurement error (bias), the impact of patient characteristics (casemix) and the play of chance also have an important impact on patient outcomes and must be taken into account before any conclusions regarding quality of care can be made. Given the often modest efficacy of medical interventions, the impact of these other factors is usually considerably greater than the impact of the processes of care, let alone the impact of variations in the quality of the processes of care. Thus, using outcomes data to obtain credible evidence with which to measure and so influence the provision of medical care is challenging. Nevertheless, as will be seen, we currently find ourselves in an era in which outcomes data are highly regarded and in which there is considerable pressure to use them in this way.

1.2 A brief history & experience of the USA in measuring quality of care

Interest in the quality of health care is not new. The Code of Hammurabi (1795 – 1750 BC) shows that the ancient Babylonians were interested in the quality of medical care, also principally in terms of its outcome. Thus, law 218 states that 'if a physician make a large incision with the operating knife, and kill him, or open a tumour with the operating knife, and cut out the eye, his hands shall be cut off', suggesting a somewhat uncompromising approach. An interest in the skill and effectiveness of physicians can also be found in other ancient cultures (Schwartz and Lurie 1990). However, the seeds of the modern era of quality assessment date back to the mid 19th century. By this time hospitals in Britain had collected patient statistics, chiefly case fatality data, for over 200 years but had made little proper use of them (Iezzoni 1996). Nonetheless, they were keenly sensitive to their unsophisticated comparison, as for example, the Glasgow Royal Infirmary whose report from 1846 stated that 'the reception of moribund cases greatly swells the number of deaths recorded in the Hospital, and very materially increases the proportionate mortality thereby producing misconceptions in the public mind...' (quoted in (Iezzoni 1996)). Florence Nightingale changed all this. She returned from the Crimean War in 1856 having used systematic record-keeping and statistical methods to show that improvements in sanitation could dramatically reduce mortality at military hospitals and determined to continue these efforts in the hospitals at home. Allied with William Farr, she set about publishing a series of influential reports, her *Notes on Hospitals*, that were packed full with comparative statistics, including observed and expected mortality rates, and ingenious statistical diagrams of her own invention, with which she helped to revolutionise attitudes to hospital

design, administration and care (Spiegelhalter 1999). She believed passionately in statistics and campaigned for all hospitals to publish a uniform set. However, she was also acutely aware of the difficulties in gathering and making sense of these data, in particular the potential for manipulation, stating 'We have known incurable cases discharged from one hospital, to which the deaths ought to have been accounted and received into another hospital, to die there in a day or two after admission, thereby lowering the mortality rate of the first at the expense of the second' (quoted in (Spiegelhalter 1999)). Sadly, her systems of publishing uniform hospital statistics were ultimately abandoned and the next major step in the measurement of quality of care was not to occur for another 50 years.

This occurred in early 20th century Boston, USA. Here, a remarkable young surgeon, Ernest Amory Codman, worked in the Massachusetts General Hospital which, like many others, reported patient outcomes at the time of discharge but made little use of the data (Neuhauser 1990). Guided by his lifelong interest, even obsession, with precision and record keeping (a keen hunter, he kept a record of every bird shot and calculated the ratio of birds shot per cartridge fired for every year of his life until 1910) Codman set about designing a better system. In 1900 he came up with the End Results System, a method by which he systematically located patients one year after discharge and brought them back for examination in order to define their disease and establish their outcome. In 1911 he set up his own hospital to implement his ideas more fully. Each patient was supplied with an end result card on which were recorded their presentation, initial diagnosis, treatments, complications, diagnosis at discharge, and outcome at one year. These cards were

summarised and an annual statement of the end results of the hospital published and specifically provided to prospective patients. Errors and adverse events were openly acknowledged and studied. By today's standards, his honesty was incredible, for example, the classification of errors included those due to lack of technical knowledge or skill, lack of surgical judgement and lack of care or equipment. His ideas proved briefly popular and even influenced the American College of Surgeons who based their initial system of hospital accreditation upon them. However, when only 89 of the 692 relevant hospitals could meet the required standards, the method was quickly abandoned. Codman himself fell out of favour and his End Results Hospital closed in 1918. The End Results System gradually fell away, no doubt because of the great efforts and honesty required by it, and with it systematic efforts to measure the quality of care again came to a halt.

Nonetheless, many of the major developments that have since occurred in the field of quality assessment have derived from the USA and an understanding of this history is useful, not least for showing how, over time, the individual elements of structure, process and outcome triad have fallen in and out of favour. Whilst interest in quality assessment in the USA no doubt reflected concern for quality of care *per se*, it is clear that the main reason for this concern derived from its free-market and somewhat unregulated medical culture. In this climate there was considerable disquiet that a significant proportion of surgical procedures were being performed without good reason (Neuhauser 1990). From 1918 onwards, the fledgling American College of Surgeons chose to use structural measures to tackle the problem. Thus, it set up a programme that specified minimum standards in terms of hospital and staff

organisation, services and training, presuming that adequate environments would spawn adequate care (Wilkin and McColl 1987). Whilst this approach undoubtedly helped to drive up the quality of hospital care in general, the method was criticised for not actually measuring the quality of care directly. As a result, in the 1940s, the focus shifted away from structural elements to attempts to measure the quality of the process of care itself (Wilkin and McColl 1987). This era was dominated by the ideas of Lembcke who promoted a 'scientific' approach in which uniform methods were used to abstract data from medical records which were then judged by explicit performance review criteria (Lembcke 1967). In line with an era where empirical proof of effective medical care was very limited, the standards used to judge quality were those that might be achieved in teaching hospitals, a rather fuzzy concept (Lembcke 1967).

Through the 1950s and 1960s, medical knowledge and resources expanded greatly and access to health care improved through the state subsidisation of medical insurance for the poor and elderly (Relman 1988). As a result, the costs of health care escalated wildly (Relman 1988). Desperate to contain the spiralling costs, the federal government intervened in the early 1970's with new methods to inspect and limit the provision of health care. Measurements of structure and process remained the tools of choice and the most important of these were the Professional Standards Review Organisations (PRSOs) whose brief was to regulate the process of care by inspecting medical records and by sanctioning certain proposed treatments *prior* to their use (Wilkin and McColl 1987). Whilst quality of care was an ostensible interest, everyone recognised that the main task was to limit activity (Sprague 2002).

Not surprisingly, clinicians disliked and distrusted the system. Furthermore, the system was methodologically poor and, as before, tended to measure processes of care without a valid link to patient outcomes (Sprague 2002; Wilkin and McColl 1987). The shortcomings of the approach were reinforced by an influential series of studies which failed to find any correlation between better performance, as measured by commonly used process-based tools, and better patient outcomes (McAuliffe 1978). With time, the concept of using measurements of the structure and process of care to indicate the quality of care fell out of favour (McAuliffe 1978; Sprague 2002; Wilkin and McColl 1987).

The early 1980's saw a new approach to cost containment with the introduction of flat rate payments for packages of care ('prospective payment'), passing the incentive to cut costs from government and insurers onto hospitals and physicians. Together with other forces at work in medicine and society (see section 1.4) a further round of interest in performance assessment followed, this time with outcomes very firmly regarded as the most important tool of measurement (Ellwood 1988). Starting in 1986, the Health Care Financing Administration (HCFA), the government body responsible for funding Medicare, published annual reports that compared mortality between hospitals across the entire nation. Others agencies joined in and eventually accreditation bodies, state governments, cities, insurers, employers, hospitals and even the media started to release information, so-called 'report cards', much of it focused on outcomes and some of it even reporting on individual practitioners (Campion and Rosenblatt 1996; Epstein 1998; Green *et al.* 1997; Iezzoni 1997a). The most well known of these is perhaps New York State's annual publication of

outcome data pertaining to individual surgeons performing coronary artery bypass graft surgery (Green and Wintfeld 1995). The quality of the data and the methods used to produce these publications varied and some of it was suspect. Process measurement continued with the PRSOs replaced by Peer Review Organisations (PROs) although their remit was little different (Dans *et al.* 1985). Thus, a confusing, disjointed and expensive morass of performance data emerged, once again much of it distrusted by the medical profession (Berwick and Wald 1990; Bindman 1999; Dans *et al.* 1985).

With time, some order has begun to appear. The HCFA terminated its publication of hospital mortality data in 1993 citing serious difficulties in their interpretation. More uniform and less adversarial approaches often emphasising evidence based process measures in addition to or in place of outcome measures have started to appear. In particular, PROs (renamed Quality Improvement Organisations) and the national agencies charged with accrediting hospitals have stopped trying to identify failing hospitals and instead now work *with* hospitals in an attempt to facilitate improvement in the quality of care at all (Campion and Rosenblatt 1996; Jencks *et al.* 2000; Jencks *et al.* 2003; Sprague 2002); and the majority of health plans in the USA now compare their performance according to the Health Plan Employer Data and Information Set (HEDIS), a uniform system which reports compliance with a number of measures of the process of care (albeit mainly in the field of prevention) and information about access and patient satisfaction (Epstein 1998). Nonetheless, a confusing proliferation of agencies and methods for measuring the quality of care persists, many continue to feel uneasy with the current measurement tools, methods

of data collection and results reporting, and consequently many remain uncertain whether the current 'quality systems' provide valid information about truly important aspects of health care and whether those measurements result in any real improvements (Bobrow 2000; Mannion and Davies 2002).

1.3 The UK perspective

Until recently, the UK has seen little of the systematic methods used in the USA to measure the quality of care. In the main, the medical profession has been left to its own devices, trust being placed in its professionalism and institutions to ensure that high quality medical care was provided. As Aneurin Bevan said 'My job is to give you all the facilities, resources, apparatus and help I can, and then to leave you alone as professional men and women to use your skill and judgement without hindrance'(quoted in (Anonymous 1998a)). An important reason for this is the different culture of health care in the UK where under rather than over provision has been the main concern (Wilkin and McColl 1987). This is not to say that medical care has not been subject to external quality reviews. Up to the 1980s there were several: for example, inspection of medical training posts by the Royal Colleges, the national confidential inquiries into maternal and peri-operative deaths, the national quality control scheme for clinical chemistry, the UK Cardiac Surgical Register, and analyses of routine data to identify mental hospitals at risk of providing poor care (English *et al.* 1984; Maxwell 1984; Yates and Davidge 1984). However, these efforts were piecemeal and voluntary rather than part of a systematic attempt to measure the quality of care across the entire NHS. Similarly, the NHS has for many

years reported a great number of process-related statistics, but primarily in terms of managerial efficiency and costs without any regard to questions of quality of care.

Things have changed very considerably over the last 15 years. In 1989, the Conservative government reformed the NHS, creating an 'internal market' in which purchasers (Health Authorities and fund-holding general practitioners) were expected to procure health care for their clients (patients and public) from providers (NHS Hospital Trusts) who in turn would compete for 'business' (Secretary of State for Health 1989). High quality information describing the provision of care was to be the 'fuel' that would power the purchasing 'engines', allowing purchasers to comprehend and influence the care they procured. Regrettably, the official information generated over this period did not match this aim but focused instead on demonstrating that 'activity' and efficiency had increased (through the flawed tools of the finished consultant episode (Clarke and McKee 1992) and the efficiency index), that waiting times or lists had reduced and that certain very simple aspects of care (such as food quality or access to single sex wards) were acceptable (Radical Statistics Health Group 1995). Questions regarding the effectiveness, appropriateness or outcome of care were simply not tackled (Radical Statistics Health Group 1995). The same period also saw the introduction of compulsory clinical audit. Although well intended, audit was bolted on to medical practice without any connection to the levers for change and became viewed as an extra chore of little benefit to be squeezed in to an already full timetable (Hopkins 1996). Consequently, whilst this period heard much talk about the need to systematically

measure and describe the quality of care, little useful data emerged and physicians became disillusioned with the entire process.

A notable exception occurred in Scotland. Since the 1960's, all hospital discharge data, cancer registrations and records of death in Scotland have been held centrally in machine readable form and capable of automatic linkage. In 1989, the Clinical Resource and Audit Group (CRAG) of the Scottish Office decided that these data should be put to use in order to help inform purchasing decisions. The pilot set of data, released in 1993, compared outcomes between Health Boards (Clinical Outcome Working Group 1993) and an expanded set, released in 1994, compared outcomes between individual Hospital Trusts (Clinical Outcomes Working Group 1994a). Annual reports followed and in many respects these data were indeed a real step forward. In particular, efforts were made to account for differences in patients' baseline characteristics and to make clear the potential impact of the play of chance. Also, clear explanatory notes were given to aid their proper interpretation. Nonetheless, partly because of the mistrustful environment into which they were released and partly because of their remaining methodological shortcomings, the CRAG outcomes data were originally received with some uncertainty (Anonymous 1993). With time, however, their publication has become less controversial and the accompanying media furore has died away (although not necessarily in the local press). No contracts have been shifted and soft evidence suggesting that the data may have helped to drive forward serivce improvements has emerged (Clinical Outcome Working Group 1998). However, until very recently, little or no rigorous

investigation of the true impact of the CRAG outcomes data has been performed (Mannion and Goddard 2003).

In 1997 a Labour government was elected and with it came further reorganisation of the NHS (Secretary of State for Scotland 1997; The Secretary of State for Health 1997). A major element was the establishment of statutory bodies to advise on best practice and to set and police minimum standards of care, including National Service Frameworks; the National Institute for Clinical Excellence and the Commission for Health Improvement in England & Wales; and the Health Technology and the Clinical Standards Boards for Scotland, more recently merged, along with CRAG, to form NHS Quality Improvement Scotland . The concept of clinical governance was introduced by which providers were expected to account for their *clinical* activities by publishing a broad range of clinical indicators. The principal measurement tool was to be clinical outcomes. In England & Wales the set of indicators, the Performance Assessment Framework, is based on Maxwell's six dimensions of the quality of care; in Scotland the CRAG indicators continue and are expanded upon. The presentation of the data in England & Wales is very similar to that in Scotland. Most recently, an incentive scheme has been added in England & Wales whereby hospitals are assigned zero to three stars according to their Performance Assessment Framework results; three star Trusts are to receive financial and administrative rewards whilst zero star Trusts are to face reorganisation and potentially external control. Thus, in contrast to the situation only a few years ago, the NHS (and especially that in England & Wales) now finds itself with a broad, complex and expanding range of instruments which purport to measure the quality of care.

Furthermore, potentially serious consequences are set to flow from their findings. The validity of these systems and their impact on the delivery of care in the NHS very much remains to be seen.

1.4 The reasons for measuring quality of care

Having established the context for the current interest in measuring quality of care, it is perhaps appropriate to ask why there should be any interest in its measurement at all. The simple answer is that measurement is needed to ensure that the standards of care are acceptable and to identify opportunities for their improvement. However, the more interesting question is why is it felt that care is not already of a uniformly high standard and that clinicians cannot simply be trusted to provide best care. There is no simple answer to this. Rather, various strands of evidence and societal shifts have joined to raise doubts about the quality of medical care in observers minds.

The first consideration is the phenomenon of unexplained variation. No matter which field of medicine is considered, studies show that patients with similar conditions treated by ostensibly similar health care providers experience different outcomes (McKee and Hunter 1995). Whilst there may sometimes be good methodological reasons for this (see section 1.6), not all the variation can be explained, and the suspicion remains that some of the differences must result from differences in care, and hence that some patients are receiving worse care than others. This is especially true when the variations in outcome are reported in sensational terms in the lay press (Anonymous 1993). However, suspicions of

variations in the processes of care are well founded. Numerous studies have shown large variations in the processes of care between health care providers between and within countries, for example, in the use of simple surgical procedures (McPherson *et al.* 1982; Wennberg *et al.* 1987), the treatment of myocardial infarction (O'Connor *et al.* 1999), the provision of cancer services (Department of Health Expert Advisory Group on Cancer 1995) and in the treatment of emergency cases (Clinical Standards Advisory Group 1995). Analyses frequently cannot account for these variations in terms of patient differences alone, in turn suggesting that they must represent either mis-use of proven interventions, variation in choice between acceptable interventions (which is to say a choice based on opinion rather than empirical fact) or variation in provision due to supply side factors (differential use of facilities simply because they exist in different amounts in different places) (Wennberg 2002). The key issue is that these differences may either have important implications for patient outcomes or, if they do not, suggest that clinicians may be providing illogical and/or unnecessary interventions. These conclusions have helped to rob physicians of some of their scientific legitimacy and encouraged purchasers and public to challenge their authority on clinical matters (Blumenthal 1994).

The second consideration is the growing realisation that medical care is potentially dangerous and that a large proportion of adverse events may be preventable. Thus, it is estimated that 3.7% of patients admitted to hospital in New York in 1984 suffered an adverse event as a result of medical care (Brennan *et al.* 1991). Of these, 13.7% suffered moderate disability with recovery within six months, 2.6% permanent total disability and a staggering 13.6% died (Brennan *et al.* 1991). In the NHS, it is

estimated that every year about 850,000 patients admitted to hospital experience an adverse event which results in harm (Department of Health 2000). These adverse events are costly, resulting in £2 billion of expenditure for the additional hospital stays (Department of Health 2000). Similarly, hospital acquired infections are estimated to cost the NHS £1 billion and yet 15% may be avoidable (Department of Health 2000). Annual NHS expenditure on settling cases of clinical negligence is £400 million (Department of Health 2000). Thus, it has become very clear to public and purchasers alike that doctors are fallible and that this fallibility is expensive. This impression has only been compounded by a series of unprecedented medical scandals, including those involving the removal and retention of human tissue at the Royal Liverpool Children's Hospital (Alder Hey), paediatric cardiac surgery at Bristol Royal Infirmary and even the mass murder of patients in Hyde, Manchester by their general practitioner (Walshe and Higgins 2002). These revelations have severely shaken the trust of the public in the medical profession and in its methods of self-regulation.

The third consideration is allied with the first and second. Recent years have witnessed great changes in the society of the UK. The public are less subservient to and less trusting of authority and at the same time more consumer orientated, demanding clear and simple information about the value of services. The public are also better informed about health care in general and about the failings of the health service in particular (Davies and Shields 1999). As such, they are empowered to ask searching questions. These societal shifts have led to a desire by the public to re-negotiate their 'contract' with the medical profession and fostered the view that the

new contract should include the establishment of systems providing evidence of good practice (Smith 1998).

The fourth consideration relates to the ever increasing cost of medical care. Over the last 30 years, this phenomenon has led virtually all modern systems of health care to limit expenditure. As noted, the main focus in the USA has been to limit medical intervention. Physicians have responded by citing fears of reduced quality to which purchasers have responded by demanding evidence of the benefits of care (Blumenthal 1996b). In the UK, with the introduction of an internal market, purchasers also became concerned to "buy right", although with our traditional lack of resources the focus of the exercise has been subtly different (Frater and Costain 1992). Nonetheless, on both sides of the Atlantic the looming question of efficiency has generated a need for data describing the quality of care in order to demonstrate value for money. Closely allied with consideration of costs, since the early 1980's the UK has also witnessed a movement to ensure accountability in all areas of public service through the use of performance indicators (Smith 1990).

The final consideration is that a climate of systematic evaluation now permeates the practice of medicine. Thus, the last decades have witnessed the rise of an empirical approach to defining medical facts (evidence based medicine) and, in particular, efforts have been made to define the efficacy of medical interventions using the randomised controlled trial (RCT) and to guide best practice through the production of systematic reviews of RCT data and clinical guidelines. By helping to identify

that which constitutes effective care this work has allowed interested parties to question more clearly than ever before whether the right thing is being done. Furthermore, coupled with modern epidemiology, the evidence based medicine movement has provided the previously missing tools with which to make valid measurements of the processes and outcome of care (Blumenthal 1996b).

1.5 The reasons for focusing on measuring outcomes

Thus, issues of unexplained variation, potential danger, lack of trust, cost-effectiveness and evidence-based medicine underpin the current powerful drive to measure the quality of care. However, a theme common to recent attempts to measure quality of care both in the USA and the UK – especially in the 1980s and 1990s - has been a profound emphasis on the importance of measuring patient outcomes and an apparent downgrading of the importance of measuring the process of care. A number of other reasons are likely to account for this, some more understandable than others.

First, there is no doubt that outcomes have a number of desirable measurement properties (Mannion and Davies 2002):

- Given that the very purpose of health care is to effect an advantageous change in a patient's health status, outcomes are rightly of prime concern to patients and public i.e. they have considerable face validity as an indicator of quality of care.
- Outcomes data are generally easy to understand.

- Outcomes are non-prescriptive, which is to say that (in theory at least) physicians may maintain the freedom to provide care as they see fit so long as the outcomes of their patients remain acceptable.
- Outcomes provide a useful summary function, acting as a clinical 'bottom line' to the often various items that make up modern packages of care, their measurement preventing the need for the collection of data on multiple, perhaps poorly understood, processes of care.
- Outcomes are useful when the question is how skilfully a procedure is performed, a question of particular relevance to surgeons (Mant and Hicks 1996).
- Outcomes are vital when the complication rate of a procedure dictates whether or not the procedure should be performed, the classic example being carotid endarterectomy where a three year risk of disabling stroke or death of much over 6% may make the operation pointless (European Carotid Surgery Trialists' Collaborative Group 1991).

To summarise the latter three points, outcomes are useful where it is uncertain that an efficacious treatment (one proven to work under experimental conditions) is also an effective treatment (one which works in the real world) (Long 1997a).

The second major reason to focus on outcomes is to do with expediency. Hospitals already collect a certain amount of data regarding patients' admissions, diagnoses, characteristics and outcome, principally in terms of case fatality. In the UK these data are relatively brief whilst in the USA they tend to be more detailed. Although the accuracy, completeness and depth of these data have been questioned (Demlo and

Campbell 1981a; McKee and James 1997a), the fundamental issue is that an unobtrusive routine system of collecting outcomes data already exists. Therefore, a routine system that attempts to measure quality of care using outcomes is currently practical and affordable in a manner that a system based on process measurement simply is not. Furthermore, it is only in the last couple of decades that we have developed both the computing power necessary to handle very large databases simply and cheaply and a wide range of validated survey instruments with which to measure important health outcomes (McDowell and Newell 1996). Thus, a further reason to focus on outcomes data is that it is only in recent times that we have really been able to do so.

The third major reason to focus on outcomes relates to our relative ignorance about the value of many of the processes of medical care. Thus, our library of RCT data remains relatively small in comparison to the immense number of health care interventions that are in use (Naylor 1995) and even where we have RCT data, it is not always clear that the findings are applicable to routine practice because the setting of the trial and participating clinicians may have been atypical, because the trial participants may have received atypical care (even placebo allocated trial participants tend to do better than non-participants) or because the patients in the trial may have been atypical (because of the use of restricted entry criteria or a low recruitment rate) (Black 1996). There are also situations in which RCT data may never be available (Black 1996). In all these situations, therefore, it can be argued that it is better to monitor patient outcomes, albeit tentatively and ideally in conjunction with some process data, rather than do nothing at all (Long 1997a).

Here, however, logic becomes fuzzy since we do not know what is or is not useful care, which brings us back to where we started from, and the object of the exercise drifts from audit to research (Davies and Crombie 1997). This is not to say such research might not be useful, but whether it should form part of a *routine* system to improve the quality of care, where the time and resources available for such analytical work are limited, is open to debate.

The final reason to focus on outcomes is that in recent times the alternative, the collection of process data, has fallen out of favour. With hindsight, however, some of the criticisms of process data, especially those made prior to the 1980s, now seem rather harsh. In particular, many of the U.S. studies that influenced the turn against process measurements have subsequently been shown to have been heavily biased against finding any association between process and outcome because of basic flaws in their measurement techniques, study designs and statistical methods (McAuliffe 1978). Furthermore, the limited state of knowledge about the value of medical interventions in the 1970s and early 1980s necessarily limited the extent to which process measures could be linked to outcomes; with more sophisticated process-based tools and better study design, process-based measures have since (Kahn *et al.* 1990) and even then (Greenfield *et al.* 1981) been shown to correlate with outcome. Lastly, dissatisfaction with the flawed PRSO and PRO systems in the USA and the domination of UK routine hospital data by the "..dreary emphasis on managerial efficiency.." (Maxwell 1984) may also have tainted the cause of process measurement.

1.6 The methodology of using outcomes to indicate quality of care

As already noted, when used to indicate the quality of care, the principles underlying the measurement of patient outcomes are those of analytical observational epidemiology. The various factors involved can helpfully be considered in the form of an equation, where V indicates variation in outcome:

$$V_{\text{OVERALL}} = V_{\text{QUALITY OF CARE}} + V_{\text{BIAS}} + V_{\text{CASEMIX}} + V_{\text{CHANCE}}$$

This formula makes it clear that outcomes can only properly be used to indicate the quality of care if it is possible to sufficiently 'cancel out' the impact of the other factors on the right hand side of the equation so that the remaining variation in outcome (or at least the greater part of it) can confidently be attributed to variation in the quality of care. The key difference from analytic epidemiology, however, is that for the purpose of measuring quality of care, measurements of patient outcomes must take place alongside routine clinical practice and should not intrude upon it. Thus many of the careful techniques of an epidemiological study cannot be used and the methods must compromise between the ideals of an academic study and that which is routinely practicable. These difficulties are outlined below.

1.6.1 Measuring an outcome

What to measure

The principal requirement of the chosen outcome is that it should be relevant i.e. important to patients and clinicians and plainly attributable to the processes of care under investigation (Giuffrida *et al.* 1999; Long 1997a). The instrument used to

measure the outcome should be valid (measure what it purports to measure), reliable (give the same outcome on repeat measurement when all other factors are held constant), responsive (show a change when there is a clinically significant change in the quality of care), acceptable (to patients in terms of language, layout and length) and practical (in terms of proven suitability to administration, analysis and communication of results) (Streiner and Norman 1989). Ideally, all these requirements should be satisfied. In practice, this is not always possible. For routine systems, where time and resources are limited, relevance, acceptability and practicality are particularly important (Long 1997b).

Over the years, the outcome most commonly used to measure quality of care has been case fatality. The advantages are clear: death is unambiguous and so relatively free from measurement bias; it is easy to collect from centrally held death certification or hospital discharge data; and, for serious disorders where death is a distinct but preventable possibility, it is undeniably important. However, most patients with potentially life-threatening disorders do not die, although their recovery may be incomplete, and many other patients have conditions with a major impact on their health but which are unlikely to cause death. An important disadvantage of case fatality data, therefore, is that they often do not provide information pertinent to the majority of patients whose care is under investigation.

To collect data pertinent to survivors, it is necessary to collect data describing the quality of survival. The World Health Organisation International Classification of

Impairments, Disabilities and Handicaps (ICDIH) provides a useful framework for this task (World Health Organisation 1980). The ICDIH describes the consequences of disease at four levels, namely: *pathology*: the damage or abnormal processes that affect an organ or system of the body; *impairment*: the loss of psychological, physiological or anatomical structure or function; *disability*: the restriction in performance of activities that results from impairment; and *handicap*: the social disadvantage that results from the interaction of impairment or disability with the individuals physical and social environment. Recently, disability has been replaced with the term 'activity limitation' and handicap with 'participation restriction' but the concepts remain the same (World Health Organisation 2001). Although not part of the ICIDH, *quality of life* is often viewed as a summary step in this hierarchy and, in terms of health, is usually taken to mean a multidimensional set of measurements reflecting the patient's physical, psychological and social functioning. Also not part of the ICIDH, *dependency* describes a state in which an individual is reliant upon others for assistance in meeting recognised needs (Wilkin 1987). It reflects the interplay between limitations in activity and the social environment and sits astride the concepts of activity and participation (Wilkin 1987).

On the whole, the outcomes of greatest importance to survivors lie in the middle (activity and dependency) and, even more so, at the higher end (participation and quality of life) of this spectrum (Roberts and Counsell 1998a). Whilst undeniably important to survivors, the problem with all these outcomes is that their great strength, their qualitative nature, means that unlike case fatality they are considerably more prone to measurement error and, because they are not routinely reported, they

are more difficult and costly to collect. With greater subjectivity and greater scope for confounding, these difficulties increase as the outcome spectrum is ascended (Duncan *et al.* 2000; Roberts and Counsell 1998a). For the same reasons, the very highest levels of outcome are also more difficult to relate to the underlying pathologies and impairments and hence to the processes of care (De Haan *et al.* 1993; Duncan *et al.* 2000), many of which primarily act at these levels. There is also a relative paucity of empirical (RCT) data linking better processes of care to better quality of life (Treurniet *et al.* 1997). For these reasons, when attempting to measure quality of care, it may be most sensible to concentrate on the collection of outcomes in the middle range of the outcome spectrum, supplemented perhaps by measurements of participation or quality of life where appropriate.

Whilst the foregoing should be the language of routine outcome surveys, their requirement for special methods of data collection means that this is often not the case. More often, routine systems must make use of those outcomes that can already be derived from routine hospital discharge data, such as readmission after discharge, complications (listed as secondary diagnoses) and place of discharge. A reasonable case can be made for each: the first two may represent failures of best care and discharge home is generally desirable. However, the convenience of these outcomes comes at the price of measurement error and difficulty in interpretation, stemming from the fact routine hospital discharge data systems were not specifically designed for their collection. Thus, readmission is sometimes a planned or a simply unavoidable event which also depends on local hospital capacity, admission thresholds and referral patterns. Readmission data are therefore only interpretable if

the proportion of readmissions that are avoidable and the proportion of patients with adverse outcomes who are readmitted are also known, which of course is not routinely the case (Gautam *et al.* 1996; Milne and Clarke 1990; Smith 1994; Thomas and Holloway 1991). Complication rates may vary between hospitals because of differences between clinicians in their vigilance and diagnostic criteria for complications, in the recording of those complications in the medical record and in the completeness of their reporting on discharge summaries; furthermore, better units may in fact have higher complication rates because they manage to keep severely ill, and hence more susceptible, patients alive (Davenport *et al.* 1996a; Silber *et al.* 1995). Discharge home is a straightforward concept and, by distinguishing those requiring some form of institutionalised care, is potentially more useful. However, it depends partly on the pre-admission home circumstances, local availability of support services and the desires and expectations of the patient and their family. Local social and cultural factors, over which a hospital has little or no control, may therefore influence return home just as much as the impact of care; measurement at the time of discharge is also problematic (see below). Thus, all of these 'convenient' outcome measures should be interpreted with caution.

Whose outcome should be measured

The sample of patients whose outcomes are measured should be truly representative of the population of patients in question. Ideally, a complete sample, derived prospectively according to rigorous and standardised methods, should be used. In the real world, routine efforts to identify patient samples are likely to have to rely on hospital data systems which are retrospective and only partially standardised. As

such, hospital samples are prone to selection biases, in particular diagnostic bias (systematic variation between hospitals in the criteria used to make a diagnosis) and coding bias (systematic variation between hospitals in the accuracy and refinement with which that diagnosis is converted into a discharge diagnostic code). If it is also necessary to obtain patients' medical records the sample may be further biased by the tendency for the medical records of deceased or complex patients to be less likely to be found (case-note retrieval bias) (Gulliford *et al.* 1991; Westgren *et al.* 1986).

When to measure

The optimum time to measure outcome represents a compromise between the provision of sufficient time for treatment effects and natural recovery to occur and a realisation that inter-current events increasingly influence the outcome as the follow up interval extends. Regardless of the interval chosen, a basic requirement is that that it should be uniform across health care providers. In many cases, this implies measurement of outcome after discharge and again argues for a special method of data collection. However, for reasons of cost, routine systems often make do with measurement of outcome at the time of discharge. Unfortunately, because length of stay varies, sometimes considerably, between hospitals, this convenient practice can result in misleading comparisons of outcomes (Jencks *et al.* 1988a). It also opens up the possibility of manipulation of outcome through manipulation of length of stay.

Measuring the outcome

A fundamental principle of medical science is that un-blinded observers make erroneous judgements about the effectiveness of interventions i.e. they systematically over or under estimate treatment effects (Noseworthy *et al.* 1994). As far from impartial observers, clinicians are unlikely to make unbiased judgements about the outcomes of patients they have treated themselves (Bilsker and Goldner 2002; Rothwell and Warlow 1995), especially if sanctions or rewards are to be applied. Ideally, therefore, clinicians should not be involved in gathering outcomes data used to measure the quality of care that they have provided. To avoid this bias, functional outcome data should be collected directly from the patients themselves and by an outside agency, implying a centralised system of follow up.

For reasons of cost, it is most likely that any such central system would have to rely on postal follow up rather than telephone or personal interview. Postal follow up has other advantages, namely, it is standardised, unobtrusive and relaxed (affording plenty of time for the respondent to consider their answers or, where necessary, for a proxy to become available) and it may also be especially suited to asking sensitive or embarrassing questions (Siemiatycki 1979). The key issue is that the response rate should be high in order to minimise the potential for non-response bias (systematic differences in outcome between responders and non-responders). The literature on the response rate to postal surveys is ambiguous, with studies reporting response rates higher (Doll *et al.* 1991; Smeeth *et al.* 2001), equal (Siemiatycki 1979) or lower (Mallinson 1998; Picavet 2001) than to personal or telephone surveys. Postal surveys are, however, prone items being left blank or being incorrectly filled,

especially amongst the elderly, reducing the effective response (Mallinson 1998; O'Mahoney *et al.* 1998; Smeeth *et al.* 2001) Simple survey instruments may reduce this bias (Brazier *et al.* 1996; Dorman *et al.* 1997a). However, the chief method to maximise response is to persistently follow non-responders. Long experience shows that this is often necessary (Dillman 1978a). A routine system which does not have this facility runs the risk of reaching invalid conclusions about patient outcomes.

1.6.2 Adjusting for casemix

Outcomes data can only be interpreted through comparison. Occasionally this may be against an absolute standard, as for example with carotid endarterectomy (see section 1.5). More often, it is against a relative standard, sense being made of an outcome by comparing it either with outcomes previously achieved at the same institution or with outcomes achieved at other institutions. In either case, for the comparison to be fair it is essential that like is compared with like, which is to say that the many patient characteristics which have an impact on outcome and which can vary over time and between settings must be taken into account. Failure to do so can result in very misleading conclusions (Aron *et al.* 1998; Green *et al.* 1990; Green *et al.* 1991; McKee and Hunter 1995; Rockall *et al.* 1995; Wen *et al.* 1995). In an experimental setting, such as a trial, variation in casemix can relatively easily be taken into account through the random allocation of patients to different treatment groups, the play of chance and other techniques ensuring a similar distribution of prognostic factors (known and unknown) in each. Randomisation has even been used to *routinely* allocate patients to different treatment groups within a single

institution for the purposes of research (Cargill *et al.* 1986). However, the routine randomisation of patients *between* institutions for the purpose of measuring quality of care would be neither sensible nor practicable. Here instead one must rely on statistical modelling techniques to control (hold constant mathematically) for differences in casemix.

The problem here is that although the statistical techniques are potentially powerful, they have their limitations. In particular, it is rarely possible to adjust fully for differences in casemix. A statistical model is an artificial construct whose power to control for casemix depends on our knowledge of the factors that influence outcome. Commonly identified prognostic variables for acute conditions include age, disease severity, the presence of comorbid disease (and its severity), physiological findings, previous medical events and 'frailty' (Orchard 1994). Often, however, many other risk factors such as gender, socio-economic status, dietary and other habits, family and social supports, psychological and cultural factors, physiological reserve, genetic predisposition and patient preferences may also be relevant. In addition to their multitude, many risk factors are difficult to define and measure and, often, their precise identity remains simply unknown. Even when identified, little may be known about the relative weights of risk factors nor the reasons they affect some patients more than others (Orchard 1994). Thus, even when excellent methods and large and detailed databases are used, statistical models based on commonly available prognostic factors may explain only a proportion of the total variation in outcome (Normand *et al.* 1996). All too frequent shortcomings in the construction and testing of statistical models may also contribute to error in their ability to adjust

for casemix (Harrell, Jr. *et al.* 1996a; Laupacis *et al.* 1997a; Wyatt and Altman 1995a). Moreover, where no single and accepted model exists, conclusions regarding hospital performance by adjusted outcome can vary quite considerably depending on the statistical model used (Iezzoni 1997b).

However, quite apart from these methodological shortcomings, in the setting of the routine measurement of quality of care, the major limitation in using statistical models to adjust outcomes for casemix is the simple difficulty in obtaining sufficient and valid data describing casemix. As noted, for reasons of cost and uniformity, routine systems often rely on the descriptive data that already happen to be available in hospital discharge returns, such as simple demographics, urgency and source of admission, principal and secondary diagnoses, and procedures performed. Whilst these data offer the possibility of making some adjustments for casemix, the method is fraught with difficulties. Most obviously, the range of data is limited and not tailored to the disorder in question, a problem of particular relevance to the description of disease severity which essentially requires clinical information. Thus, in comparison with the use of more relevant data, the ability to account for variation in outcomes using routinely coded data is considerably reduced (Green *et al.* 1990; Green *et al.* 1991; Hannan *et al.* 1997; Hartz.A.J. and Kuhn 1994). Similarly, the ICD coding system with which primary and secondary diagnoses are reported fails to distinguish between mild and severe cases and, because the system lacks clear clinical definitions, it allows for the differential assignment of risk factors between hospitals in otherwise identical patients (Iezzoni 1990). The standard of clinical information supplied to coding clerks and the guidelines used to derive coded data

may also lead to incompleteness and inaccuracy in routinely reported risk factor information (Iezzoni 1990). Variation between hospitals in the degree of error in these data can significantly bias the comparison of adjusted outcomes, the impact potentially rivalling that of variation in quality of care (Green and Wintfeld 1993).

Particular biases relate to secondary diagnosis data. Secondary diagnoses are those conditions other than the primary condition which were also present on admission (comorbidities) or which developed during the admission (complications) and which affected the management of the patient (Anonymous 1990). Comorbidities and complications are reported together and yet when adjusting for casemix it is only legitimate to adjust for comorbidities since the development of complications may, in part, reflect the quality of care provided. Failure to exclude complications from adjustments for casemix may therefore lead to 'over-adjustment', falsely compensating and hence obscuring poorly performing hospitals (where complications are presumably more likely to have occurred) (Blumberg 1991; Hannan *et al.* 1997; Shapiro *et al.* 1994). The problem is that it is often not easy to distinguish comorbid conditions from complications, even on detailed review (Jencks *et al.* 1988a). Furthermore, to ignore all complications may be simplistic and lead to under-adjustment given that some do indeed occur simply because of the severity of the patient's condition (Shapiro *et al.* 1994). Thus, it is very difficult to define which secondary diagnoses should be selected to make adjustments for casemix. Added to this difficulty is the fact that physicians and coding clerks tend to report complications in preference to comorbid conditions when a patient dies (Green and Wintfeld 1993; Iezzoni *et al.* 1992; Jencks *et al.* 1988a; Romano and Mark 1994).

This sometimes leads to a paradoxical association of comorbidities with survival and hence biased adjustment for casemix (Iezzoni *et al.* 1992; Jencks *et al.* 1988a). The use of risk factor data derived from clinical investigations (e.g. CT head imaging after stroke) may also be problematic when there is appreciable variation in the use of that investigation between centres. Here, variation in the presence of the risk factor (e.g. haemorrhagic stroke) may then simply reflect variation in the use of the test. Lastly, if reimbursement for hospital care depends on routinely reported patient characteristics (e.g. if greater payments are made for treating sicker patients) there may be a tendency for hospitals to over-report the presence of certain comorbidities (Hsia *et al.* 1988).

A number of strategies might be employed to reduce these shortcomings, for example, altering discharge returns to allow the reporting of a greater number of secondary diagnoses, allowing comorbidities to be reported separately from complications, and improving and monitoring the quality of routine discharge information. Nonetheless, it is clear that credible adjustments for casemix are only likely to result if specific and detailed clinical data are collected for the task. Particularly in the USA, efforts to obtain such data have focused on the clinical data contained within the medical record. However, this approach also has its important shortcomings. First, the completeness and accuracy of information may vary between hospitals; second, the collection of data is prone to inter-observer error, an important consideration since a large number of non-medical staff are usually required for the task (Iezzoni 1994); and third, the approach is extremely expensive. For example, in the early 1990s, the total annual cost for a system in California was

estimated to be $61.2 million (Iezzoni 1997b). For this reason alone, few of the states which originally collected or even considered collecting casemix data from the medical record have continued with it (Iezzoni 1997b).

A prospective approach would clearly be the best way to gather clinically relevant casemix information but a routine system would require the widespread co-operation of physicians and also new and perhaps costly methods of data collection. The method is sometimes used in specific circumstances, as for example, in the reporting of outcomes after cardiac surgery in New York State (Green and Wintfeld 1995) and in the United Kingdom (Fine *et al.* 2003). However, even here data collection problems may occur either in the terms of incompleteness and reduced reliability (Fine *et al.* 2003) or bias resulting from physicians reporting information that is then used to judge their own performance (Green and Wintfeld 1995).

1.6.3 The play of chance

In some ways it can be argued that one need not account for the play of chance when comparing hospital outcomes data. After all, the sample of patients for whom the outcome is being reported is usually the entire population of patients at each hospital and hence the question of random error in the estimate of outcome should not arise. However, this policy ignores the fact that the play of chance is a very important arbiter of the outcome of medical care, a phenomenon well illustrated by a comparison of risk adjusted case fatality between UK neonatal intensive care units in which, without any feedback of data, hospital rankings fluctuated very widely from

one year to the next (Parry *et al.* 1998a). Therefore, in order to take account of the impact of the play of chance, it is widely agreed that, for the purposes of measuring quality of care, each hospital population should be viewed as a sample drawn from an underlying population of potential patients (Thomas and Hofer 1999). Thus, statistical tools can and should be used to quantify the play of chance (Goldstein and Spiegelhalter 1996). In keeping with most medical research, a p value of 0.05 is widely used, although stricter p values are sometimes used to account for multiple comparisons. For graphical comparisons of unadjusted data, the point estimates of outcome are usually accompanied by confidence limits that indicate hospitals with outcomes significantly different to the mean; for risk adjusted data, confidence limits are used to indicate whether the point estimates of the observed outcomes are significantly different to the expected outcomes. Hospitals with outcomes that are significantly different to the average or expected value are termed 'outliers'. By definition, outliers are unlikely to have experienced their outcomes because of chance and, hence (so the theory goes) provided proper account has been taken of bias and casemix, have a high probability of having provided exceptionally good or exceptionally bad care.

Regrettably, there are two problems with this approach. The first is the simple question of numbers. Even for important medical conditions such as stroke or myocardial infarction, in statistical terms most hospitals admit relatively few patients per condition per year (usually a few hundred) amongst whom fewer still experience an outcome event. Thus, the degree of uncertainty in any estimate of outcome over one year is high and so the ability to confidently differentiate between hospitals with

different outcomes is limited and the rankings of hospitals by adjusted outcomes are frequently unstable, changes over time being more likely to reflect regression to the mean rather than to changes in the quality of care (Goldstein and Spiegelhalter 1996; Jencks *et al.* 1988b; Luft and Hunt 1986; Marshall and Spiegelhalter 1998; Parry *et al.* 1998a; Rothwell 2000).

The second problem, intimately connected to the first, relates to the need to stipulate a cut-off within the outcome distribution above which one may assume that a poor outcome is unlikely to be due to chance. To properly understand this problem, it is helpful to view the overall outcome distribution as being made up of two over-lapping bell curves: a large curve derived from hospitals with acceptable quality of care and a smaller curve, shifted toward the right-hand tail of the large curve, derived from hospitals with poor quality of care. The 'quality cut-off' (usually the upper 97.5^{th} percentile) within the overall distribution bisects the two constituent bell curves, leaving, in terms of poor quality of care, true and false positives on one side and true and false negatives on the other. The key questions, therefore, are what proportion of hospitals beyond the 'quality cut-off' truly provide poor quality of care (i.e. what is the positive predictive value of outlier status) and what proportion of poor quality hospitals remain hidden within the body of the *overall* outcome distribution (i.e. what is the sensitivity of using outcomes to identify low quality of care)?

Somewhat disappointingly, rigorous statistical simulation studies suggest that even under ideal conditions (*total* adjustment for casemix and large differences in potentially avoidable mortality between average and poor quality hospitals) for common medical conditions the impact of random variation is such that the positive predictive value and sensitivity of outcomes data is surprisingly low (Hartz.A.J. *et al.* 1997; Hofer and Hayward 1996; Thomas and Hofer 1999; Zalkind and Eastaugh 1997). For example, in a comparison of hospitals each with a sample size of 200 patients, for common medical conditions outlier status was estimated to have only a 38% positive predictive value and a 12% sensitivity for poor quality of care (Thomas and Hofer 1999). Greater precision in the estimates of outcome and hence more accurate predictions of quality of care result if sample sizes are increased. However, even with hospital sample sizes of 900, simulation studies suggest that for common medical conditions the positive predictive value of outlier status may only reach 32% and sensitivity only 68% (Thomas and Hofer 1999). Furthermore, for many hospitals, reaching these sample sizes is difficult and requires the summation of outcomes over time or across diagnoses. Summation over time is the most sensible approach but if data must be aggregated over several years the resulting outcomes data may be of little relevance to the current situation (Jencks *et al.* 1988b). The alternative, summation over diagnoses, allows the data to be more timely but, given the poor correlation of the outcomes of patients with common medical disorders admitted to the same institution, the method would very likely simply trade random error for systematic error and so obscure any problems in individual conditions (Jencks *et al.* 1988b; Rosenthal *et al.* 1998).

1.6.4 Interpretation, publication & unintended consequences

The foregoing makes it abundantly clear that, especially when current data collection systems are used, the collection of patient outcomes entails considerable deviations from the best practices of analytical epidemiology. As a result, our ability to 'cancel out' the impact of measurement error, variation in casemix and the play of chance is limited. Hence, even at their best, comparisons of outcomes result only in hypotheses about relative performance which then require further investigation by direct measurements of the process of care. The rationale for using outcomes data therefore is one of efficiency, the method in theory limiting the need for more complex investigation. The successful application of this strategy depends very much on the outcomes data having a high positive predictive value and a high sensitivity for poor quality of care, which as noted, may be a problem. Unless the positive predictive value is high, considerable resource will be expended on unnecessary investigations of the process of care and the morale of staff and the confidence of the public in their local hospital (to which, in the UK, there is usually no alternative) will in many cases be needlessly damaged. Unless sensitivity is high, the ability of the entire exercise to influence and improve the overall quality of care will be low. A further and related difficulty of using outcomes as a screening tool is the simplistic dichotomisation that results, hospitals and physicians being divided into groups either likely or unlikely to have deficiencies in care; the so-called hunt for "bad apples". The problem with the bad apples approach is that it ignores the reality that the process of care is unlikely to be perfect at *any* provider and hence that all have the scope to improve their services in one way or another. A danger in using

outcomes data, therefore, is that by allowing the majority of hospitals to 'shelter' in the body of the outcomes distribution, this important fact may become obscured.

These issues are brought sharply into focus by the decision whether or not to publish comparisons of outcome data. Those against publication suggest that the limitations of comparative outcomes data will not be understood outside of scientific circles and hence that erroneous conclusions about quality of care will be drawn (Goldstein and Myers 1996; Marshall *et al.* 2000) and that those with poor outcomes will be more likely to try to explain away their results (e.g. 'our patients are sicker', 'poor quality data', etc.) rather than to explore the data in an attempt to improve quality, as would occur if the data were quietly fed back to providers (Anderson 1999; Thomson *et al.* 1997). Proponents, on the other hand, suggest that publication of outcomes data may help the public and purchasers to make rational health care choices and, for fear of coming bottom in any league table, may spur all providers into improving services, something which would not occur if poor results could be quietly ignored (Leatherman and McCarthy 1999; Marshall *et al.* 2000). Empirical studies conducted in the USA suggest that, in fact, consumers, clinicians and purchasers are little interested in and/or do not understand published performance data but that hospitals and health care provider organisations are interested in them, are able to respond, and that in some cases health improvements have resulted (Leatherman and McCarthy 1999; Marshall *et al.* 2000). That said, recent work (related to the CRAG outcomes data) suggests that even hospitals take little note of official, published performance data in the UK (Mannion and Goddard 2003). In truth, however, too little time has elapsed and too few studies have been performed to draw definitive

conclusions (Leatherman and McCarthy 1999; Marshall *et al.* 2000). In the meantime, given the current demand for public disclosure of information in nearly all walks of life, it is likely that the public disclosure of comparative outcomes data will continue.

An important consideration in the use of performance indicators is that, in addition to any planned benefits, their use may have a number of unintended consequences, especially if the indicator data are made public (Smith 1995; Tymms and Wiggins 2000). Reflecting the understandable desire of clinicians to avoid coming last in any league table, the unintended consequences most relevant to the publication of outcomes are the deliberate manipulation of data in order to appear to have better outcomes than is truly the case (a practice known as 'gaming') and the deliberate limitation of clinical practice to cases where a good outcome is likely. Examples of both have been mooted in relation to the publication of mortality rates after cardiac surgery in New York State, USA (Green and Wintfeld 1995). Thus, between 1989 and 1991, the reporting system went from being a non-controversial internal activity to a league table of individual surgeons outcomes published in the lay media. Over the same period, the reported incidence of patient risk factors jumped considerably. For instance, the average incidence of chronic obstructive pulmonary disease (COPD) jumped from 7% in 1989 to 12% in 1990 and then to 17% in 1991; at some hospitals the increase in the incidence of COPD over this period was almost unbelievable, up from 2% to 53%. Whilst some of this increase may have resulted from a change in risk factor definition, there is a strong suspicion that some may have resulted from a deliberate attempt by some clinicians to portray their patients as

'sicker' than was truly the case and hence to spuriously lower their adjusted case fatality rates (Green and Wintfeld 1995). At the same time, lay and medical journals suggested that, as result of the publication of outcomes data, high risk patients were more likely to be refused cardiac surgery in New York State or were being referred for cardiac surgery in another state (Green and Wintfeld 1995; Omoigui *et al.* 1996). Interestingly, however, neither of these suspicions were confirmed by later more rigorous investigation (Peterson *et al.* 1998), and with stability in their definition and the introduction of a system to inspect their accuracy, the large fluctuations in the prevalence of risk factors has much reduced (Chassin *et al.* 1996).

1.6.5 Summary

This review makes plain the considerable methodological hurdles and limitations inherent in using outcomes to indicate the quality of health care. On an optimistic note, it is clear that many of the methodological barriers might be overcome with better system design, in particular the establishment of a dedicated system to identify patients and their baseline characteristics, to inspect the validity of that data, to collect important outcome data in a valid manner, and to adjust for casemix using powerful and externally validated statistical models. Of course, each of these 'simple' steps is challenging and the establishment of such a system would undoubtedly be expensive. Unfortunately, the question of low numbers and hence of uncertainty in those comparisons is less easily solved and, when used as the sole indicator of quality, the focus of outcomes on "bad apples" is likely to remain. Summarised in a recent non-systematic review (Thomas and Hofer 1998), studies

that have directly addressed the question of whether adjusted outcome data do indeed indicate quality of care have given mixed results (Thomas and Hofer 1998). Thus, uncertainty, tinged with quite some doubt, continues to reign over the use of outcomes to indicate the quality of care provided by hospitals, even more so that provided by individual clinicians. Despite this, the pressure to routinely measure and compare patient outcomes continues and, as noted, such data are currently routinely published for hospitals in the UK. Prominent among these, both in Scotland and in England & Wales, are data describing outcome after stroke. A consideration of the reasons for this state of affairs and of the literature relating to the use of outcomes to measure the quality of stroke care forms the subject of the chapter that follows.

Chapter Two: Stroke and the quality of stroke care

2.1 The importance of stroke

Stroke is of considerable importance. It is estimated that it is the second commonest cause of death and the sixth commonest cause of premature disability (including mortality) in the world (Murray and Lopez 1997a; Murray and Lopez 1997b). In the UK, it is estimated that there are 125,000 new strokes every year and that of men and women aged 45 years, nearly 25% and 20%, respectively, can expect to suffer a stroke if they live to be 85 (Bonita 2002). The majority of patients survive their stroke (Thorvaldsen *et al.* 1995) and hence the major impact of stroke is felt in terms of chronic disability and impairment (Wolfe 2000). At one year after stroke about one third of survivors remain dependent on other people in order to perform simple activities of daily life (Bamford *et al.* 1990a). Stroke is the second commonest cause of severe disability amongst adults living in private households in the UK (Martin *et al.* 1988) and is a leading cause of dementia, depression, epilepsy and falls. The most common stroke related residual impairments are in the areas of cognition, speech and use of the lower limbs (Wolfe 2000). The economic burden of stroke is also large and accounts for nearly 6% of total annual NHS and Social Services expenditure (about £2.3 billion) (Anonymous 1998b). The indirect costs of stroke (loss to the workforce, sickness benefits, early retirement) are also substantial. However, the great burden of stroke falls mainly on the elderly with about three quarters of all strokes occurring in those over 65 years of age and about half in those over 75 years of age (Bamford *et al.* 1988). Given that the current rapid increase in

the elderly population, and hence the incidence of stroke, these burdens are only set to increase, especially in the area of acute care (Malmgren *et al.* 1989).

2.2 The provision of stroke care

Despite its clear importance, stroke has only recently been viewed as a significant priority by most policy makers and health service professionals (King's Fund Consensus Conference 1988). According to the report of a consensus conference held in 1988, services for stroke patients were "...haphazard, fragmented and poorly tailored to patients' needs.." and treatment was hampered by "...a striking lack of convincing data on the effectiveness of widely used medical, psychological, and specific rehabilitative treatments." (King's Fund Consensus Conference 1988). Since then, however, interest in stroke has grown considerably, beginning with the publication of a series of studies which detailed its epidemiology (Bamford *et al.* 1988) and which showed the effectiveness of strategies of primary (Atrial Fibrillation Investigators 1994; Collins *et al.* 1990) and secondary prevention (North American Symptomatic Carotid Endarterectomy Trial Collaborators 1991; European Carotid Surgery Trialists' Collaborative Group 1991; Antiplatelet Trialists' Collaboration 1988; Collins *et al.* 1990). By far the most important advance came in the early 1990s when evidence of an effective *treatment* emerged, a systematic review showing that when compared to conventional ward care, organised stroke unit care resulted in a 21% reduction in the odds of death at one year (Langhorne *et al.* 1993). An expanded review subsequently confirmed this benefit and showed that it was not simply at the expense of an increased proportion of disabled survivors, stroke unit

care resulting also in a 31% reduction in the odds of death or dependency and a 25% reduction in the odds of death or institutionalisation (Stroke Unit Trialists' Collaboration 1997a). This period also witnessed considerable interest in the use of thrombolysis to treat acute ischaemic stroke, although only one trial (The National Institute for Neurological Disorders and Stroke rt-PA Stroke Study Group 1995) of several reported any significant benefit. By the mid 1990s – the period of the SOP - trials investigating the use of anti-thrombotic drugs in acute ischaemic stroke were actively recruiting (International Stroke Trial Collaborative Group. 1997) and data suggesting the potential value of specific interventions such as physiotherapy (Langhorne *et al.* 1996) and the prevention of dysphagia (Barer 1989; Smithard *et al.* 1996) were beginning to emerge.

In the last few years further progress has been made. In particular, studies have started to un-pick the 'black-box' of stroke unit care and plausible reasons for its efficacy have started to emerge (including preservation of physiological homeostasis and early mobilisation in the acute phase; aggressive prevention and intervention for complications; specialisation and education of staff; the formal co-ordination of activities toward common goals; and the education and involvement of carers) although the precise impact of individual aspects of organised stroke care remain to be formally tested or reported (Evans *et al.* 2001; Hoening *et al.* 2001; Indredavik *et al.* 1999; Langhorne *et al.* 1999; Langhorne and Dennis 1998; Langhorne and Pollock 2002; Pound *et al.* 1999; Stroke Unit Trialists' Collaboration 1997b). Aspirin has been shown to have a small but real impact on acute ischaemic stroke, resulting in nine fewer serious vascular events (non-fatal myocardial infarction, non-

fatal stroke or death from vascular cause) per 1000 treated (Antithrombotic Trialists' Collaboration 2002); the limited role of acute anticoagulation has become clear (International Stroke Trial Collaborative Group 1997); the potentially large impact of thrombolysis on ischaemic stroke has emerged with an estimated 44 extra patients alive and independent per 1000 treated when used within six hours of onset and 126 extra patients alive and independent per 1000 treated when used within three hours of onset (Wardlaw *et al.* 2003); and there has been considerable expansion in the field of secondary prevention, particularly in relation to antiplatelet drugs (Antithrombotic Trialists' Collaboration 2002), blood pressure and cholesterol reduction (Heart Protection Study Collaborative Group 2002; PROGRESS Collaborative Group 2001; The Heart Outcome Prevention Evaluation Study Investigators 2000) and strategies for dealing with carotid stenosis (Brown 2003; Rothwell *et al.* 2003). Throughout this period, current best practice has been periodically distilled into authoritative clinical guidelines both within the UK and at a European level (King's Fund Consensus Conference 1988; Royal College of Physicians 1989; Scottish Intercollegiate Guidelines Network (SIGN) 1997a; Scottish Intercollegiate Guidelines Network (SIGN) 1997b; Scottish Intercollegiate Guidelines Network (SIGN) 1997c; Scottish Intercollegiate Guidelines Network (SIGN) 1998; The Intercollegiate Working Party for Stroke 2002; Aboderin and Venables 1996).

Thus, in the space of about 15 years, our understanding of and ability to intervene in stroke has undoubtedly improved. However, the implementation of this new knowledge throughout the UK has been patchy, a situation highlighted in 1995 by a survey of consultant physicians (who routinely treated patients with stroke) which

showed that only 44% had access to a specialised stroke rehabilitation unit, that only 17% worked in a hospital where they could identify a physician with a special interest in stroke, that 20% had no access to non-invasive carotid imaging and that about one third of patients were admitted to hospitals without on-site CT scanning facilities (Lindley *et al.* 1995). Whilst a number of surveys have subsequently documented improvements in these areas, particularly in the provision of specialised rehabilitation, the fact remains that there are unacceptably large variations between hospitals and regions in the UK in access to and use of stroke services (Ebrahim and Redfern 1999; Rudd *et al.* 1999; Rudd *et al.* 2001a; Rudd *et al.* 2001b). Thus, for example, in the period 2001-2, only 27% of patients admitted to hospital with stroke spent more than half of their hospital stay on a stroke unit (an improvement from 18% in 1998) and the documentation of care remained poor with only about two thirds having a record of a swallowing assessment or a visual field examination recorded in the medical record (Rudd and Pearson 2002). Disturbingly, there are also data to suggest that the outcomes of patients admitted to hospital with a stroke in the UK, both in terms of mortality and morbidity, may be worse than that of patients in continental Europe (Wolfe *et al.* 1999) and elsewhere in the world (Weir *et al.* 2001) and that outcome after stroke also varies between regions within the UK (Rudd *et al.* 2001a).

2.3 Measuring the quality of stroke care

2.3.1 The situation in the UK

Thus, classic pressures (disease importance, high costs, availability of effective interventions, and known variation in processes and outcome) have focused attention on the need to measure the quality of stroke care in the UK. To this can be added the governments decision to make stroke a priority for disease prevention and treatment under its Health of the Nation programme (Dennis and Warlow 1991) and the introduction of the purchaser-provider split to the NHS in the 1990s, with the implication that purchasing decisions would be guided by (and hence the need for) data describing clinical activities (Miles *et al.* 1995). A number of attempts to measure the quality of stroke care have followed.

Clinicians have been primarily interested in structure and process data and focused initially on the development of a credible and reliable measurement instrument, the Royal College of Physicians Stroke Audit Package (RCPSAP). The RCPSAP is an audit form comprising 60 evidence and consensus based criteria and is used to assess the completeness of the medical record (and hence, in theory, the quality of care) in relation to key items of stroke assessment and intervention (Anonymous 1994; Gompertz *et al.* 1994). Initially the RCPSAP was designed for *ad hoc* use. More recently, an up-dated and expanded version (the Intercollegiate Stroke Audit Package) that also incorporates criteria relating to the organisation of care has been used in three national audits of stroke care in England, Wales and Northern Ireland (Rudd *et al.* 1999; Rudd *et al.* 2001b; Rudd and Pearson 2002). The national clinical

guidelines for stroke in England & Wales and in Scotland recommend the periodic collection of data describing the process of care for the purpose of quality improvement (Scottish Intercollegiate Guidelines Network (SIGN) 1997a; The Intercollegiate Working Party for Stroke 2000). In Scotland, a systematic survey of the structure and some elements of the process of stroke care has been performed in order to identify any deficiencies in relation to clinical guideline statements of best practice (Anonymous 2003a); and surveys of structure and process have also been carried out in the remainder of the UK under the auspices of the Stroke Association (Ebrahim and Redfern 1999; Lindley *et al.* 1995) and more recently by the British Association of Stroke Physicians (Rodgers *et al.* 2003).

However, over the last 10 years government and purchasers have been primarily interested in measuring the quality of stroke care using comparisons of outcome, in particular using outcomes derived from hospital discharge returns. As noted, this may simply reflect the desire to make best use of data collection systems that already happen to be in place. However, it is also arguable that outcomes might actually be peculiarly suited to measuring the quality of stroke care because of the great number of small and repeated processes of care that are involved (that might otherwise have to be measured individually) and because the key element of stroke care, its formal organisation on a stroke unit, is still only partially understood and hence difficult to define for the purposes of measurement (Langhorne and Dennis 1998).

Scotland was the first country within the UK to publish comparisons of routinely collected stroke outcomes between hospitals in 1994 with the release of data describing 30 day mortality and the proportion of patients discharged to their home address within 56 days of admission from there (Clinical Outcomes Working Group 1994a). Similar data followed in England & Wales in 1999 (NHS Executive 1999). Regular reports have followed with outcomes aggregated over three year periods in Scotland and annually in England & Wales; in each case 95% confidence intervals are given. However, very limited attempts have been made to control for variation in casemix. For hospital level comparisons adjustment is made only for age, sex and socio-economic status; in England & Wales comparisons are then also stratified by hospital type. Thus, a major shortcoming of the published data is that the comparisons of outcome are not adjusted for more important prognostic variables such as pre-stroke functional status and living arrangements nor for the severity of the stroke, factors which can vary between hospitals, even those of similar type, (Rudd *et al.* 2001a) presumably because of differences in the characteristics of the populations served and because of differences in local referral patterns and admission thresholds. Furthermore, neither system directly reports disability or dependency, the outcomes of greatest importance to survivors of stroke. Thus, coupled to anxieties regarding the quality of routine hospital discharge data and the difficulties in allowing for chance, the value of the stroke outcomes data published in the UK remains in some doubt.

2.3.2 Brief review of previous work

The literature describing the use of *observational, risk-adjusted outcomes data* to indicate the quality of stroke care gives only limited guidance. On balance, it seems that a relationship between better adjusted outcome and better stroke care can be discerned at a population level. The first large study in this field was carried out by the RAND corporation in the USA which, in a sample of 2824 stroke patients, showed a significant association between 30 day mortality, adjusted for casemix, and the quality of the process of care as judged by explicit review of the medical record (Kahn *et al.* 1990). More recently a study of 6,666 patients from 149 Veterans Affairs hospitals in the USA showed that, after adjusting for casemix, populations treated in better staffed, co-ordinated and resourced rehabilitation units, which provided better staff education and which dealt with higher volumes of patients had better outcomes in terms of discharge to community accommodation (Hoening *et al.* 2001). Two large studies from Sweden have shown that patients who were independent prior to their stroke, conscious on admission and treated on a stroke unit were significantly less likely to be dead or institutionalised at three months, and more likely to be independent in activities of daily living at two and a half years, than similar patients treated on the general wards (Glader *et al.* 2001; Stegmayr *et al.* 1999). Two smaller but more detailed studies of 128 highly selected stroke patients from 11 Veterans Affairs hospitals in the USA showed a positive relationship between improved structure and process of care (using systematically developed and tested measurement instruments) and better adjusted functional outcome at six months (Duncan *et al.* 2002a; Hoening *et al.* 2002). Studies investigating the provision of stroke care by physician type (neurologist vs. others) (Mitchell *et al.*

1996) and by method of reimbursement (Kramer *et al.* 2000) have also shown that patients with greater access to specialist investigations and/or rehabilitation facilities had better adjusted outcomes in terms of case fatality and long term institutionalisation, respectively, and a study comparing American academic centres found that units with vascular neurologists, with guidelines limiting the use of thrombolysis and (just not significantly) with a mobile stroke team had lower adjusted in-hospital case fatality (Gillum and Johnston 2001).

However, not all studies have been supportive. The same study of American academic centres failed to find any relationship between lower adjusted case fatality and greater access to stroke unit care (Gillum and Johnston 2001); a large American study of 3,611 patients (based on routinely collected data) failed to find higher case fatality, after adjusting for casemix, in those judged to have poor quality of care (as defined by Peer Review Organisation review criteria) (Thomas *et al.* 1993); and a small study found no relationship between adjusted case fatality and greater access to specialist rehabilitation (Retchin *et al.* 1997). On the whole, however, the negative or ambiguous studies are less credible than the positive ones because they generally make less detailed or relevant adjustment for casemix (Gillum and Johnston 2001; Retchin *et al.* 1997; Thomas *et al.* 1993) (only one positive study does the same (Mitchell *et al.* 1996)); because of a reliance on limited and potentially dubious measurements of the process of care (Thomas *et al.* 1993); and because of the use of an insensitive outcome measure (case fatality rather than disability when the difference in care was access to specialised rehabilitation) (Retchin *et al.* 1997).

Studies in which the hospital is the unit of analysis represent a much sterner and more realistic test of the validity of using outcomes to indicate the quality of stroke care. Thus, whilst population level analyses ask the question of whether in general, or (more often) in selected groups, better outcomes are associated with better care, hospital level analyses ask the more difficult question of whether this relationship can be discerned when relatively small and heterogeneous populations (a substantial proportion of whom may derive little benefit from the measured treatments) are compared. At this level, the literature relating to the use of observational, risk-adjusted outcomes to indicate the quality of stroke care is much less encouraging.

Of three studies which have made a comparison between hospitals with and without a stroke unit and which have accounted for clinically important casemix, only one (a before and after study' by Duncan *et al.* (1995)) showed that stroke unit care resulted in a substantial reduction in mortality. Oddly, however, it did not show any benefit in terms of improved functional status. The two other studies (one a 'before and after' study (Davenport *et al.* 1996b), the other a comparison between two districts (Gompertz *et al.* 1995)) failed to find a significant relationship between stroke unit care and better outcome in terms of case fatality, functional status or discharge home. The same is also true of a more recent study which measured the process of care at three hospitals using the RCPSAP and adjusted for clinically important casemix (McNaughton *et al.* 2003). However, all these studies can be criticised because of their small size, both in terms of the numbers of hospitals studied and numbers of patients at each, a failing especially pertinent to the studies conducted by Duncan *et al.* (1995) and by McNaughton *et al.* (2003) in which the hospital samples included

fewer than 100 patients. Nonetheless, studies involving larger numbers of hospitals and patients have reached similarly negative conclusions. Thus, an American study failed to find any difference in a hospital-wide measure of the structure of care between 62 hospitals classified as either high or low outliers in terms of adjusted 30 day stroke mortality (Jessee and Schranz 1990); another study of 21 American hospitals failed to find any correlation between adjusted case fatality and the proportion of cases judged to have poor stroke care (as defined by Peer Review Organisation review criteria) (Thomas *et al.* 1993); a study of 12 European hospitals could not explain the residual differences in survival and dependency, after adjusting for important differences in casemix, by further adjustment for measured differences in the process of care (length of stay, type of bed used, access to brain imaging) (Wolfe *et al.* 1999); and another study of 12 American hospitals (six high and six low quality 'outliers') could not find any relationship between in-hospital adjusted mortality and the quality of the process of stroke care when measured by explicit process review criteria (although it did find some some rather soft evidence suggesting fewer potentially preventable deaths at low mortality outlying hospitals) (Dubois *et al.* 1987). Although not strictly a hospital level analysis, a recent study which amalgamated consecutive series of about 40 patients admitted to hospitals within the same region within the UK also failed to find any relationship between the regional mean scores for the structure and process of care (measured using the Intercollegiate Stroke Audit Package (ISAP)) and outcome (in terms of 30 day mortality, length of stay, disability or institutionalisation) after adjusting for important differences in casemix (Rudd *et al.* 2001a). However, the interpretation of these larger studies is also difficult for various reasons, including: failure to make

stroke specific measurements of the structure of care (Jessee and Schranz 1990); limited measurements of the process of stroke care (Thomas *et al.* 1993); unreliable collection of the process data (Thomas *et al.* 1993); limited adjustment for casemix (Jessee and Schranz 1990; Thomas *et al.* 1993) or, where more detailed data were potentially available, failure to properly adjust for stroke severity (Rudd *et al.* 2001a); failure to relate adjusted outcomes to a hospital level measure of the quality of stroke care (Wolfe *et al.* 1999); by the use of small numbers (the study by Dubois *et al.* (1987) involved only 106 patients); by the use of amalgamated rather than true hospital samples (Dubois *et al.* 1987; Rudd *et al.* 2001a); and by a general reliance on retrospectively collected data (Jessee and Schranz 1990; Thomas *et al.* 1993; Dubois *et al.* 1987; Rudd *et al.* 2001a).

In summary, the literature appears to show that, provided they are appropriate and adjusted for clinically important casemix, outcomes data show a *generally* positive relationship with the quality of the structure and process of stroke care, especially if patients unlikely to benefit from care are excluded from the analysis. Importantly, however, it also appears that there is no body of evidence to show that this positive relationship can be reliably discerned when the outcomes of hospitals are compared. That said, because of the real and potential shortcomings of the hospital level studies performed to date, methodological reasons may account for this fact. In particular, it is clear that the literature continues to lack a study that tests the strategy of using outcomes to indicate the quality of stroke care in a large numbers of hospitals, in which adjustments are made for important differences in casemix *and* in which important aspects of the structure and process of stroke care are measured.

2.4 The Stroke Outcomes Project

The Stroke Outcomes Project (SOP), the study upon which this thesis is based, was designed in the early 1990s in response to the decision by the Scottish Office to commence publication of routinely collected stroke outcomes data and the concern that, because of their crudeness, these data might mislead rather than inform. Extra impetus for the study derived from the medical and media attention that greeted the first set of outcomes data (Anonymous 1993; News at Ten, Independent Television News 1994) and the suggestion that, in the internal market then in operation in the NHS, purchasing decisions might be based upon them. More generally, the limited state of the literature (especially in the mid 1990s) relating to the use of outcomes to indicate the quality of hospital stroke services demanded that higher quality hospital level studies be performed.

The fundamental premise of the SOP was that if routinely collected outcomes data were indeed to be the sole official measure of the quality of stroke care in Scotland, then it was imperative that an improved method of measuring and comparing stroke outcomes be developed and tested. In particular, the SOP wished to explore two key improvements: the possibility of routinely measuring functional status in addition to case fatality (in order to report an outcome of direct relevance to survivors) and the possibility of routinely adjusting outcomes for clinically important prognostic variables. The methods used to determine the feasibility and validity of this 'ideal' yet routine system are detailed in the chapter that follows.

Chapter Three. The Stroke Outcomes Project

3.1 Introduction

As noted, the purpose of the Stroke Outcomes Project (SOP) was to investigate an improved yet still routinely applicable system of measuring outcome after stroke. Specifically, the SOP aimed to investigate:

1. Whether it is feasible to collect valid measures of functional status in addition to case fatality and to adjust these outcomes for important differences in casemix.
2. Whether these outcomes can then be used to indicate the quality of the structure and process of stroke care.

Of course, in order to truly define the value the proposed system (i.e. to define the sensitivity and positive predictive value with which it might identify hospitals with poor quality of care) we would have had to study tens of hospitals. Regrettably, the limited resources available meant that, in fact, we could study only five. Right from the outset, therefore, it is important to appreciate that the SOP is akin to a pilot study, capable of exploring the validity of our proposed system but unable to provide a definitive answer.

The SOP involved the collection of data using both retrospective and prospective means. The methods were complex and can be divided into five inter-related steps:

1. the identification of patients admitted to hospital with an acute stroke;
2. the collection of data describing casemix;

3. the collection of data describing the structure and process of stroke care.

4. the collection of data describing important outcomes after stroke;

5. adjustment of the outcomes for important differences in casemix.

This chapter describes the methods underlying each step.

3.2 Study setting

Hospital characteristics

The five participating hospitals were all acute NHS Hospital Trusts in the central belt of Scotland. They took part on the understanding that they would remain anonymous and therefore I will refer to them as Hospitals A, B, C, D and E. We deliberately selected these hospitals to represent a range of settings, the different hospitals varying in terms of urban or rural catchment area, admission threshold, teaching or district general hospital status and in the degree of organisation of stroke care. The broad differences between the Trusts at the start of the study are summarised in Table 3.1. The hospitals differed in the number and location of their constituent parts. A summary of the make up of each hospital is given in Table 3.2.

Outcome after stroke 1990-1993

We selected the hospitals such that their 30 day case fatalities spanned the range of the 1994 publication of stroke outcome indicators (Clinical Outcomes Working Group 1994b) (Figure 3.1) in which adjustment was made for age, sex and ICD cerebrovascular disease code. Hospital D had the lowest adjusted 30 day case

fatality (21%) in Scotland while Hospitals A and E had the highest (38%) and second highest (37%), respectively. The case fatalities of Hospital B and C were intermediate to these cases (30% and 31%, respectively) and were not significantly different to the Scottish average.

Study period

We identified patients admitted between 1 August 1995 and 31 July 1997. The study therefore reports comparisons of stroke outcomes collected over two rather than three years, the period used in the publication of stroke outcome data in Scotland. The study period represents a compromise between the need to collect sufficient numbers to minimise the impact of chance and the resources available for the project.

3.3 Identifying potential cases of stroke using routine methods

Currently, the only method that allows the routine identification of patients admitted to hospital with a stroke in Scotland is the inspection of hospital discharge data. The study method of identifying stroke patients is therefore based on these data.

3.3.1 Routinely collected hospital discharge data in Scotland

Consultant episodes and the SMR1

Scottish hospital discharge data are reported as 'consultant episodes', a term used to describe an occasion on which an in-patient is under the care of a individual

consultant. In theory, after every consultant episode, a Scottish Morbidity Record (SMR) is submitted to the Information and Statistics Division (ISD) of the Scottish National Health Service. In practice, SMRs are generally only produced when patients are discharged from hospital or transferred to another department within the same hospital. The production of SMRs when patients are transferred between consultants within the same department is unusual. There are several different types of SMR form; that relevant to stroke is the SMR1. The SMR system was upgraded during the study and hospitals switched to the new version at different times between April 1996 to April 1997 i.e. starting eight months into our study period (Anonymous 1995).

Cerebrovascular Disease Codes

The principal diagnosis given on an SMR1 is defined as the main condition managed or investigated during that episode of care. Diagnoses are coded according to the International Classification of Disease (ICD). The ninth revision (ICD 9) was replaced with the updated version, ICD 10, on the 1 April 1996. The codes used to describe the different categories of cerebrovascular disease are given in Table 3.3. Many, but not all, of the ICD 9 codes map to those of ICD 10.

Diagnostic codes on the SMR1 were assigned by trained coding staff in each hospital. Codes were assigned on the basis of clinical information provided by clinicians on hand-written or typed discharge summaries or, in some cases, using data recorded in the medical record itself. The limited amount of data pertinent to

stroke patients that are collected on the SMR1 in addition to the primary diagnosis are shown in Table 3.4.

Hospital stays

The term 'hospital stay' is used to refer to a continuous period of in-patient treatment. Patients are frequently transferred from the care of one consultant to another during their hospital stay and, as noted, in these situations, two or more SMR1s may be generated. In particular, for stroke, one SMR1 may be generated when a patient is transferred from the admitting to the rehabilitation team and a second when the patient is discharged from rehabilitation. More complicated admissions may result in a number of SMR1 forms. Difficulties with SMR1 data are that those pertaining to the same hospital stay do not always give the same principal diagnosis and the delay between admission and the generation of an SMR1, and between a first and any subsequent SMR1, may sometimes be considerable. Also, ISD link SMR1 returns to provide a record of completed hospital stays only periodically.

3.4.2 Identifying patients for the SOP

Our choice of method for identifying patients for the SOP was principally driven by our aim of identifying functional outcome in survivors at six months after admission. This requirement meant that our system of identifying patients had to be *fast* and *on-going* to allow us to follow patients up by this time.

Eligible cases

We based our system of patient identification on the assumption that the great majority of patients admitted with an acute stroke would have a cerebrovascular disease code listed as the primary diagnosis on at least one SMR1. We therefore set up a link with ISD such that each month we were sent an electronic file of every SMR1 generated by the study hospitals that had *any* cerebrovascular disease code listed as the primary discharge diagnosis. We modified our collection of SMR1 data after the first six months in light of an analysis of the accuracy these data (see section 4.3.2). We continued to collect monthly SMR1 data until March 1998 in order not to miss patients admitted towards the end of the recruitment period.

Study database

We entered the SMR1 data onto a computerised database (Microsoft FoxPro 2.6). Since it was necessary that we receive SMR1s with the minimum of delay, it was impossible for ISD to link them to identify hospital stays when those stays consisted of two or more consultant episodes. We therefore established our own system of identifying hospital stays, automatically linking SMR1s if the patient identification data matched (surname and initial, sex and post code *all* the same) and if the dates of admission and discharge overlapped or were immediately adjacent. This system also identified duplicate SMR1s (SMR1s referring to the same consultant episode for the same patient) and kept only one on the study database.

Where there was a gap of more than one day between SMR1 dates belonging to the same patient, our system could not identify whether the SMR1s referred to one stay (reported as separate consultant episodes) or to two separate stays. Similarly, our system could not identify instances where patient identification data varied, in error, between SMR1 returns e.g. a single hospital stay reported by two SMR1s, each with a slightly different date of birth. During data collection, therefore, our database listed a mixture of full stays (all SMR1s correctly linked, copies deleted), partial stays (unlinked SMR1s) or duplicate stays (SMR1s we could not identify as a copy). The practical result of this was that some patients on the database appeared to have been admitted more times than was the case. We therefore put systems in place to prevent multiple audits and follow up of the same hospital stay (see below). Once data collection was complete and prior to analysis, we linked or deleted these 'extra' stays as appropriate to create a final database that described only full hospital stays.

Thus, using routine methods, we established a cohort of patients at each hospital who *might* have had an acute stroke. The next task was to identify which of these cases had truly had an acute stroke and to collect data describing their casemix and the quality of the care provided.

3.4 Baseline data collection: background

3.4.1 Research Assistant

We employed a single research assistant to collect data describing diagnosis, baseline characteristics and the process of care from the medical record. Previously, she had worked as registered hospital nurse for ten years and, in the four years immediately before the study, she was employed as a senior audit assistant and coding clerk. She therefore had considerable clinical experience and a good understanding of the requirements and difficulties of retrospective review. She was given detailed instructions in the study methods of data abstraction and, before starting, audited 30 medical records as a test exercise. Her interpretation of the recorded data and the accuracy with which she collected the study variables were checked and any errors in her technique corrected.

3.4.2 Data abstraction

Casemix and process of care data extracted from the medical record refer to complete hospital stays. For the purpose of baseline data collection, we defined a complete hospital stay as the interval between admission and death in hospital or discharge from in-patient rehabilitation. Patients transferred to long term care facilities within hospital were treated as if discharged. Where patients were clearly undergoing rehabilitation whilst housed in a ward designated for long term care, we defined the end of a hospital stay as the date after which active rehabilitation ceased.

We inspected all entries to the medical record (medical, nursing and, if available, therapist) pertaining to the stay in question and recorded abstracted data on a standard proforma (Appendix 1). Our research assistant resolved any difficulties through discussion with medically qualified staff. We checked the repeatability of some of our data collection by means of an inter-rater study (see chapter 5).

3.4.3 Assignment of patients to hospitals

We assigned patients and their outcomes to the hospital (identified by audit the medical record) to which they were *first* admitted. This rule allowed us to deal with patients transferred between hospitals. For example, for a patient transferred from Hospital A to Hospital D, we assigned the measurements of the process and outcome of care to Hospital A alone. If patients were first admitted to a *non*-study hospital and then transferred to a study hospital, we excluded them from the study altogether. It is recognised that patients transferred between study hospitals are difficult to classify (Jollis and Romano 1998). Fortunately, other than for patients with SAH (who are excluded from analysis, see below) transfer between hospitals for stroke is unusual. Any bias resulting from our method of assigning patients to hospitals should therefore be small.

3.5 Baseline data collection: diagnosis and casemix

3.5.1 Diagnosis of acute stroke

Acute stroke

We decided not to use the formal WHO criteria to define cases of acute stroke (Asplund *et al.* 1988a). We reasoned that its use would have entailed our non medically qualified research assistant using the limited data available in the medical record to second guess the diagnosis assigned by the clinicians who cared for the patient, a difficult and potentially inaccurate task (Asplund *et al.* 1988a). Instead, we used a pragmatic definition, namely whether a diagnosis of acute stroke was noted in the medical record. In the great majority of cases, a final diagnosis of an acute stroke assigned by a hospital physician, especially if supported by a CT head scan, is likely to be correct (Ferro *et al.* 1998a; Kothari *et al.* 1995; von Arbin *et al.* 1981a). We accepted the diagnosis of acute stroke only if it was the diagnosis assigned by the most senior physician whose opinion was recorded and if the delay between the onset of the stroke and admission to hospital was 30 days or less. We included patients admitted to hospital with stroke and those whose stroke occurred after admission for another disorder. If a patient suffered a second stroke after admission, this was ignored. We used a 30 day period to define acute stroke because the prognostic models we intended to use to adjust for casemix were developed and validated using a very similar definition (see section 3.9.1). Furthermore, patients presenting after this period represent a cohort of survivors whose outcome may be little influenced by the measured items of stroke care.

We assigned each hospital stay to one of four diagnostic categories:

1. Acute Stroke — Stroke verified by audit
2. Ineligible stroke — Stroke verified by audit BUT over 30 days between onset and admission and/or transferred from a non-study hospital
3. Not stroke — Non stroke disorder defined by audit
4. Unknown — Insufficient data in the medical record to establish a diagnosis

'Not stroke' included re-admissions (after an admission for acute stroke) if the patient was discharged for longer than 24 hours. We recorded the principal reason for admission in all cases of 'not stroke' based on the most senior recorded opinion.

Pathological subtype of acute stroke

We identified the pathological subtype of an acute stroke if the result of a CT head scan, lumbar puncture or autopsy was recorded in the medical record. We assigned acute strokes to one of four pathological categories: ischaemic stroke; haemorrhagic stroke; subarachnoid haemorrhage (SAH); not known. If the CT head scan was reported as normal, or if no bleed was reported, we assumed that the patient had had an ischaemic stroke regardless of the interval between the onset of the stroke and the scan. Where we could not identify the pathological type of stroke we assumed that it was not SAH.

Study patients: acute, non-SAH stroke

We restricted the principal analyses in the SOP to patients with an acute, non-SAH stroke. Patients with SAH are excluded because the presentation, treatment and outcome of this condition is distinct to that of other pathological subtypes of stroke. Moreover, our methods of adjusting outcome for casemix are not applicable to patients with SAH.

3.5.2 Collection of casemix data

All hospital stays

We collected limited identification and demographic data for all hospital stays (Table 3.5). We calculated age as the number of decimalised years from birth until admission to hospital (i.e. 70 years 6 months = 70.5 years).

Social deprivation

We derived the socio-economic status from the Carstairs deprivation score which, in turn, is derived from the postcode of the patient's usual place of residence (McCloone 1994). The Carstairs score is based on four markers of social deprivation (overcrowding, male unemployment, low social class for the economically active head of the household and car ownership) and is calculated for each postcode sector (the area described by the first four digits of the code) in Scotland using small area census data. On the basis of the Carstairs score, each postcode sector is assigned a deprivation category ranging from one (most affluent) to seven (most deprived).

Acute stroke

In cases of acute stroke, we also obtained data describing patients' characteristics prior to admission and the severity of their condition at presentation. The variables selected for this purpose were all recognised or potentially important predictors of outcome after stroke (Counsell and Dennis 2001) and which, when the SOP was designed, were undergoing investigation by our department (see Table 3.6).

We abstracted data describing severity at presentation from entries relating to the first 24 hours of admission. If a patient's condition varied during this 24 hour period, we used the worst scenario except for the measurement of systolic blood pressure, where we used the most representative reading for that day. The only exception to the 24 hour time interval for data describing baseline severity was for urinary incontinence which was noted over the first seven days of admission. If the medical and nursing notes disagreed with regard to data describing severity at presentation, we collected the data in the medical notes (except for urinary incontinence, in which case we preferred the nursing notes).

3.6 Baseline data collection: the structure and process of stroke care

We wished to base our audit of the structure and process of stroke care primarily upon items that had been shown to improve outcome after stroke in large randomised controlled trials (RCTs) and systematic reviews or items whose importance was derived from authoritative statements of best practice that were in existence or in preparation at the time of the study i.e. the mid 1990s (Aboderin and Venables

1996b; King's Fund Consensus Conference 1988b; Royal College of Physicians 1989b; Scottish Intercollegiate Guidelines Network (SIGN) 1997c).

Regrettably, our use of review criteria based on unequivocal evidence was limited. In the mid 1990s only four interventions were supported by evidence from large RCTs or systematic reviews: organised stroke care (Langhorne *et al.* 1993b), secondary prevention with aspirin and anticoagulation after ischaemic stroke (Antiplatelet Trialists' Collaboration 1994; Atrial Fibrillation Investigators 1994), and carotid endarterectomy in patients with a recent, non-disabling ischaemic stroke and a critical stenosis of the ipsilateral carotid artery (North American Symptomatic Carotid Endarterectomy Trial Collaborators 1991; European Carotid Surgery Trialists' Collaborative Group 1991). Anticoagulation and carotid endarterectomy are applicable to only a small proportion of patients and hence are only likely to have a small impact on hospital outcome (Warlow *et al.* 1996). We therefore elected not to collect information about carotid endarterectomy at all. We did collect information about the use of anticoagulation but with the intention of reporting it in conjunction with the use of aspirin i.e. the overall use of anti-thrombotic medication.

The value of CT brain imaging in acute stroke is supported only by opinion and observational data (Donnan 1992; Wardlaw 1994). However, it is an important diagnostic tool and a pre-requisite to rational secondary prevention using anti-thrombotic drugs and carotid endarterectomy, and is best performed within a few days of stroke onset (Bogousslavsky *et al.* 1991; Dennis *et al.* 1987). We therefore

included measures of the access to and speed of CT brain imaging amongst our principal measures of the structure and process of care.

A further problem was the difficulty of defining organised stroke care. We wished to define organised care by its supposed fundamental characteristics: the provision of care in a geographically defined unit; the co-ordination of care by regular and formal multidisciplinary team (MDT) meetings; the specialisation and expertise of staff in stroke; and the provision of education and training for staff and carers in stroke (Langhorne *et al.* 1993b; Langhorne 1995; Stroke Unit Trialists' Collaboration 1997c). However, auditing even these limited parameters was difficult. In particular, there is no accepted definition of a SRU or expertise in stroke care and potentially important aspects of care such as the proper co-ordination of rehabilitation, staff education and carer involvement in rehabilitation cannot be measured by auditing the medical record. The criteria we used to assess the organisation of care are therefore somewhat simplistic.

3.6.1 The structure of stroke care

We collected data to describe the organisation of care, the wards and the availability of key non-medical staff involved in providing stroke care, access to diagnostic facilities, outpatient facilities and other miscellaneous items (Table 3.7). We performed two surveys of the structure of stroke care, the first in December 1995 and the second in April 1997. The first survey was performed by the research assistant only; the second by both myself and the research assistant. At each hospital, we

collected data by interviewing the study liaison physician, the therapists involved in providing stroke care and qualified nursing staff on the wards that usually cared for patients with stroke. Our audit assistant identified the wards to which stroke patients were usually admitted and we ascertained the number of beds per ward from information held by the medical records departments of each hospital. In addition, because our research assistant visited most hospitals each month, we were able to keep abreast of important changes in the structure of care at other times during the study period by informal contacts with hospital staff.

3.6.2 The process of stroke care

Data collected in all cases of acute stroke

Table 3.8 shows the data describing the process of care collected for each patient. Our principal measures of the process of care were: the provision of organised stroke care; access to CT brain imaging; and the prescription of an anti-thrombotic drug on discharge. We also measured the use of certain other items of acute stroke care whose value was less certain or unknown (at the time of the SOP): use of subcutaneous heparin, administration of parenteral fluids and insertion of a urinary catheter.

For all items, we collected the date on which their provision/use was *first* recorded. When there was no record of an item of care in the medical record, our audit of the process of care did not differentiate between circumstances where this appeared appropriate (e.g. in the presence of a recognised contraindication) and where it

appeared to represent a failure of care. The methods employed in an attempt to take account of patients' characteristics are detailed in chapter eight.

In each case, we defined the provision of organised stroke care by three criteria:

1. Whether there was a record that a MDT meeting had been held *at any time*. We accepted that a MDT meeting had been held if the notes clearly stated that this was so, or, where this was not explicit, if the opinions/management decisions of members of three or more therapeutic disciplines (e.g. physician, speech therapist and physiotherapist) were recorded at the same note entry.
2. Whether the patient was admitted to a geographically defined stroke unit, and, if so, the dates of admission and discharge.
3. Whether the patient was discharged from the care of a physician with a special interest in stroke. A physician with a special interest in stroke was defined as one with specific responsibility for providing the stroke service in that hospital.

Where there was a geographically defined stroke unit we also collected further information about selected aspects of their usual practice by discussion with members of staff on that unit (Table 3.8).

The Royal College of Physicians Stroke Audit Package (RCPSAP)

We audited the completeness of the *physicians* entries to the medical record (i.e. ignoring any nursing or therapists sections of the medical record) in relation to the 60 key items of history, examination, investigation and management described in the original version of the Royal College of Physicians Stroke Audit Package (RCPSAP; see Appendix 2) (Gompertz *et al.* 1994b; Hancock *et al.* 1997; McNaughton 1996). In addition, we gave two RCPSAP criteria individual consideration. First, the swallowing question, which enquires whether attention to the ability to swallow safely has been documented within 24 hours of admission. Observational data and expert opinion hold that this is an important part of the initial assessment because it appears to reduce the incidence of aspiration pneumonia and hence improves outcome after stroke (Barer 1989; Smithard *et al.* 1996). Second, the 'clinical diagnosis' question, which enquires whether the initial clerking finishes with a summary detailing the patient's neurological deficit, the likely site of their cerebral lesion and the relevant risk factors. Such a summary demonstrates stroke specific thinking on the part of the admitting physician and is held by the developers of the RCPSAP to be the most important step in developing a management plan (Anonymous 1994).

Unlike our main process of care audit criteria, the RCPSAP audit criteria allow for the omission of items of care in the presence of certain common contraindications e.g. failing to record an assessment of the ability to swallow is disregarded if the patient is documented as being unconscious. Such appropriate omissions of items of care are recorded as 'no, but...' whereas inappropriate omissions are recorded as

'no'. Thus, compliance with each item of care is calculated as the proportion of patients in whom the item of care was relevant (those with yes or no responses) in whom the standard was achieved (compliance = number of yes responses/ number of yes and no responses). A similar method can be used to calculate a compliance score for the RCPSAP as a whole. We collected RCPSAP data in a representative sample of about 100 patients with acute stroke at each hospital. We derived this sample by selecting every nth case on the audit list at each hospital, n varying to ensure a similar number of cases audited across hospitals. Where a case was selected to have RCPSAP data collected but did not have an acute stroke, the case immediately below on the audit list was audited.

Patients transferred between hospitals

For patients transferred between study hospitals, we audited the process of care at both hospitals and assigned all process data to the first hospital. However, if a patient was transferred from a study hospital to a non-study hospital (as was possible at Hospitals B and C) we were *unable* to audit the medical records at the transfer hospital. Our measurements of the process of care at the transfer hospital were therefore limited to the information contained in any discharge letter resulting from that care. Again, however, other than for patients with SAH (who are excluded from analysis) transfer between hospitals for stroke is unusual.

3.7 Outcome: definitions and timing

3.7.1 *Outcome measures*

In addition to case fatality, we measured functional outcome in survivors in terms of the need for help from others in activities of daily living (dependency) and in terms of place of residence. We reported dependency and residence in conjunction with case fatality, giving the combined outcomes of 'death or dependency' and being 'alive and at home'. These measures are important, relevant to patients, easy to interpret, influenced by intervention and are recommended for and widely used in trials of acute treatment and secondary prevention after stroke (Duncan *et al.* 2000; Roberts and Counsell 1998b; van Gijn 1992). They are also capable of being adjusted for important differences in casemix by our prognostic models (see below). Case fatality and 'alive and at home' also approximate the two measurements of outcome after stroke that are currently published in Scotland.

Although collected, I have not reported data describing quality of life collected using the Euroquol, a measure of health related quality of life after stroke (Dorman *et al.* 1997b). This instrument measures physical function, depression and anxiety and the presence of pain. In the absence of a validated method of adjusting the Euroquol data for casemix, their comparison between hospitals is unlikely to be useful for measuring the quality of care. However, the data are useful descriptors of patients *per se* and I have therefore used them to compare the casemix of responders and non-responders to our outcome questionnaire (see chapter six).

3.7.2 Definition of outcome states

We defined case fatality as death from any cause.

We measured dependency using the dependency question of the 'simple questions' originally used in the International Stroke Trial (International Stroke Trial Collaborative Group 1997b; Lindley *et al.* 1994a). This question asks: do you need help from anybody with everyday activities? Our main reason for using it in the SOP was the assumption that, in a routine system where repeated follow up of non-responders might not be possible, a simple and brief outcome measure would be most likely to encourage a high level of response. The absence of any clinically important bias when the question is answered by proxy is a further practical advantage (Lindley *et al.* 1994a).

A potential drawback to using the simple dependency question is that it has only been demonstrated to have *reasonable* validity and reliability (Dennis *et al.* 1997b; Dennis *et al.* 1997a; Kay *et al.* 1997; Lindley *et al.* 1994a). Although perfectly adequate for comparisons of outcome between very large groups of patients (thousands in each), its use may reduce our ability to identify differences in outcome between smaller groups, such as between the hospitals in this study. In the last eight months, we also collected functional status using the Modified Rankin Scale (MRS) (Bamford *et al.* 1989) in order to compare its findings with those of the simple dependency question. This scale is widely used as a measure of functional status after stroke and allows a finer classification than the simple dependency question:

scores 0, 1 and 2 indicate functional independence at increasing levels of difficulty while scores 3, 4 and 5 indicate increasing levels of functional dependency (see Appendix 3).

We determined place of residence by means of a specific question. This question asked: 'where are you staying now?' and gave five alternatives: same place as before illness; in hospital; with family; in a nursing or residential home; other (asked to specify). For the purpose of the SOP, we defined 'at home' to mean living somewhere other than a hospital or a residential/nursing home. Thus, we considered patients living at the same place as before their illness or with family to be 'at home'. For those living in an 'other' place, we inspected the accompanying details and assigned their place of residence accordingly. Where patients indicated that they lived in sheltered accommodation we defined them to be 'at home'.

3.7.2 Timing of outcome assessment

We measured outcome six months after admission, a period widely regarded as sufficient to allow for natural recovery and the impact of treatment to become apparent and that is recommended for use in trials of treatments of stroke (Dombovy *et al.* 1987; Duncan *et al.* 2000; Jorgensen *et al.* 1995; Stroke Unit Trialists' Collaboration. 1997b; Wade and Hewer 1987). Equally important in choosing a six month follow up were the practicalities of our system of identifying patients. SMR1s are frequently only completed after the patient is discharged, and in all cases up to four weeks could elapse before ISD passed the SMR1 on to us. ISD estimate that six

months need to elapse after the end of a given period in order for 99% of SMR1s to be completed and returned (Clinical Outcomes Working Group 1999a). Thus, using our system of identifying cases, the first point at which it is actually possible to send a follow up after a fixed interval to the majority of cases is approximately six months after admission. Lastly, we followed patients at six months because our prognostic models are able to make predictions of outcome at this time.

*3.7.3 **Definition of outcomes measures** (as used in main analyses)*

Case fatality: all analyses of case fatality are based on linked survival data (see section 3.8.1), with six month case fatality (from any cause) calculated at 183 days after the first day of the complete hospital stay, taking the admission date obtained by auditing the medical record as baseline.

Death and dependency: we identified case fatality at six months as above. In cases alive at six months, we preferentially used the simple question to identify dependency; if the dependency question was left blank but the MRS was filled, we defined patients with a score of 3, 4 or 5 as dependent.

Alive and at home: we identified case fatality at six months as above. In cases alive at six months, we used the response to the question 'where are you staying now' (see 3.7.1) to define place of residence.

All survey data pertaining to patients alive at six months, regardless of when the questionnaire was actually filled, were defined as 'six month outcome data'. This rule allowed us to include returns that were late either because there was delay in identifying their stay (late SMR1 data) or because of delay in mailing, filling or returning the questionnaire.

3.8 Outcome: methods of follow up

We used separate methods to identify those patients who had died and to collect data describing the quality of survival in patients alive at six months. The fundamental principle underlying each was that the method had to be feasible on a routine basis i.e. capable of being used on a large scale (national) and on-going basis.

3.8.1 Case fatality

Unlike other parts of the UK in the 1990s, Scotland was fortunate in that it already possessed an accurate and simple system for routinely identifying survival at a fixed time after admission to hospital. This was possible because ISD were able to link SMR1 data to the Registrar General's death certificate records and do so with a minimum of 99% accuracy (Clinical Outcomes Working Group 1999a). Given that a practical and reliable routine system of identifying death was already in place, there seemed little point in developing an alternative. The draw back to the method is that linkage to death certification data is currently performed only periodically. We obtained linked fatality data from ISD and therefore had to wait until ISD had

identified six month survival status for all patients in the study. These data were not available till 1999, nearly two years after the last patient was recruited. Therefore, although the linked data provided a definitive and routine measurement of survival at this time, we could not use them to guide our system of following up survivors. In the future much more rapid linkage may be possible (Clinical Outcomes Working Group 1999a).

3.8.2 Following up survivors

We set out to test a centralised and semi-automated system to follow survivors using a postal questionnaire. Our outcome questionnaire is shown in Appendix 4.

We designed our system to overcome two practical difficulties to routine postal follow up in Scotland. First, SMR1 data do not give the address the patient was discharged to or even the address from which they were admitted; instead, they simply give the *post code* of the patient's residence *before* they were admitted. Second, we had no way of knowing if patients had died after discharge and wished to avoid mailing questionnaires to recently bereaved relatives. However, SMR1 data do give the address and sometimes the name of the patient's general practitioner (GP) (the latter is an optional field) *prior* to admission to hospital and, we reasoned, the GP was highly likely to know whether the patient was alive and if so, to know their current address. We therefore designed a two stage system of follow up that relied on the GP to act as our 'middle-man'.

First, we checked all hospital stays listed on our database to determine whether the patient had died in hospital. Then, in all patients where the SMR1 discharge code showed that the patient was discharged alive, we sent a follow up pack to the pre-admission GP surgery, naming the relevant GP whenever possible. The follow up pack consisted of two parts, one for the GP and one for the patient.

The 'GP pack'

This included an explanatory letter, a single sided response form and a reply paid envelope (Appendix 5). We requested that the GP tell us if and when the patient had died or, if the patient was still alive, to forward our 'patient pack' to them. We also requested that the GP let us know if the patient was no longer on their list. The explanatory letter made it clear that the patient's consultant had let us know about the hospital stay, that the purpose of the survey was to collect data to measure the quality of hospital care and that any data collected would be confidential.

The 'patient pack'

This was provided as an envelope (left unsealed so the GP could inspect the contents) with the patient's name and a postage stamp already attached; the surgery simply had to add the patient's address and post the envelope on. The pack included a brief explanatory letter (Appendix 6), a questionnaire and a reply paid return envelope. To encourage response, we tried to make the explanatory letter personal by using the patient's name and by mentioning that the consultant in charge of their

care. The letter also made it clear that the information requested would be kept confidential and would help in trying to improve care at their local hospital.

Posting follow up packs

We posted the follow up pack to the GP three weeks before the six month follow up date to allow sufficient time for the questionnaire to reach the patient. Follow ups were generated and sent on a weekly basis. In cases where we first received SMR1 data more than six months after the patient was admitted to hospital, follow up was sent in the next weekly batch .

Minimising the number of follow ups sent

Because our database sometimes listed erroneous extra hospital stays (see section 3.4.2), there was the potential for some patients to be sent more than one questionnaire. The same applied to patients who genuinely had two or more hospital stays for acute stroke in quick succession. In both cases, to reduce workload and avoid antagonising GPs, we established systems to minimise the number of follow ups. Thus, only one follow up was sent and the same outcome data used if the ISD admission dates of two stays were within 30 days of one another or if the follow up form for the first stay was completed within the 30 day period before the follow up form for the second stay was due to be sent. To further reduce follow up, when we became aware that a hospital stay was due to an 'ineligible stroke' or a 'not stroke' disorder before follow up was due, no follow up was sent.

Non-standard GP responses

Soon after starting follow up, it became apparent that GPs were making two common mistakes. First, instead of forwarding the patient pack, some would reply that it was 'fine for us to contact' the patient or simply confirm that the patient was alive. Second, some GPs would reply telling us the patient's current address. We reasoned that these same mistakes would likely occur in a real system and altered our follow up system accordingly: where GPs gave their permission for us to mail a questionnaire or confirmed the patient was alive, we sent back another follow up pack with a letter reiterating that we could not follow the patient ourselves; where we were provided with the current address, we forwarded a patient follow up pack ourselves.

Returned questionnaires

We defined the date on which follow up actually occurred as the date the questionnaire was completed. We did not attempt to contact GPs or patients in cases of non-response for the first 14 months of the study. For the final 10 months we modified our methods of follow up in an attempt to improve response (see chapter 6)

3.9 Outcome: adjusting for casemix

3.9.1 The study prognostic models

We adjusted the comparisons of outcome between hospitals for variation in casemix using multiple logistic regression. To do so, we applied three simple prognostic

models (the 'study models') recently developed within our own department (Table 3.10). The development of these models and their predictive properties are described in detail elsewhere (Counsell *et al.* 2002; Counsell 1998). Briefly, the models were rigorously derived according to established guidelines (Laupacis *et al.* 1997b; Wasson *et al.* 1985; Wyatt and Altman 1995b) using a training data set taken from the Oxfordshire Community Stroke Project (OCSP), an unselected cohort of patients with first ever stroke (Bamford *et al.* 1988). Each model uses the same six simple clinical variables, each of which can be collected during a single visit to the patient's bedside. The external validity of the models was tested in two independent cohorts: a community-based cohort of 538 patients with first ever stroke (Anderson *et al.* 1993; Ricci *et al.* 1991) and a hospital-based cohort of 1330 patients with first and recurrent stroke (derived from the prospective stroke register of the Western General Hospital, Edinburgh). These data sets provided a stern test of the external validity of the models by assessing their accuracy across time, place and patient type (Braitman and Davidoff 1996; Harrell, Jr. *et al.* 1996b; Justice *et al.* 1999; Laupacis *et al.* 1997b; Wyatt and Altman 1995b). The models describing survival and independent survival were found to have good external validity and the model for 'alive and at home' reasonable external validity (Counsell *et al.* 2002; Counsell 1998).

3.9.2 Applying the study models to the SOP data set

Two terms, discrimination and calibration, are used to describe the accuracy of the predictions of a logistic regression model (Harrell, Jr. *et al.* 1996a; Justice *et al.* 1999). Discrimination refers to the ability of a model to differentiate between

individuals who do and do not experience an outcome whilst calibration refers to the accuracy with which a model predicts outcomes for *groups* of patients. In the comparison of adjusted outcome between hospitals, it is most important that a model is well calibrated.

The calibration of an externally derived prognostic model can be improved by customising the coefficients of its predictive variables to the dataset under investigation i.e. by building a new model in the study data set, forcing in exactly the same predictive variables as used in the original model (Moreno and Apolone 1997; Shu *et al.* 1996). Ideally, I wished to use customised models to adjust the comparisons of all three outcomes in the SOP. Unfortunately, a customised model is only likely to be valid if there is complete ascertainment of outcome, which, in the SOP, was only likely to be the case for death. For the outcomes collected by postal survey (dependency and residence) it was inevitable that there would be losses to follow up. I therefore used a customised version of the study model only for the outcome of case fatality. For the outcomes of 'death or dependency' and 'alive & at home' I used the models in their original format.

3.9.3 Measures of adjusted outcome

I used two complimentary measures of adjusted outcome.

1. p values and odds ratios

I used logistic regression analysis to estimate the statistical significance of any overall variation in outcome between hospitals and to describe the outcomes of the hospitals relative to one another. To do so, I entered a hospital term as a categorical variable, with Hospital A as the reference category, into each regression analysis. I took the significance of the hospital term overall (*not* of the individual comparisons A vs. B, A vs. C, etc.) to indicate the significance of any variation in outcome between hospitals. I calculated odds ratios to describe the outcomes of the hospitals relative to Hospital A. I derived the odds ratios and their 95% confidence intervals using the formula

$$e^{(\textit{Logistic Coefficient of each hospital} + / - 1.96 \times \textit{Standard Error})}$$

For unadjusted outcomes, I performed a logistic regression analysis with the hospital term as the only covariate. For adjusted case fatality, I performed a logistic regression analysis with the hospital term entered in addition to the covariates of the customised study model. For adjusted 'death or dependency' and 'alive and at home' I first calculated a linear predictor value for each patient in order to preserve the predictive properties of the study model (the method is given in appendix 7) (Moreno and Apolone 1997; Shu *et al.* 1996). I then performed a logistic regression analysis with the linear predictor and the hospital term as the only covariates. All logistic regression analyses were performed using SPSS version 9.0

2. The W score method

The W score method provides a measure of adjusted outcome in absolute terms. W scores are easy to interpret and have been used in previous studies investigating variations in mortality between centres and over time (Parry *et al.* 1998b; Smith *et al.* 1990; Yates *et al.* 1992). The W score measures the difference between the observed and the predicted number of outcome events per 100 patients treated *within* a hospital. The W score is calculated using the formula

$$w = (o - p) / (n/100)$$

where o is the observed number of events, p the predicted number of events and n the total number of patients per hospital. For example, if 500 patients are treated, 150 deaths observed and 100 deaths are predicted, then the W score is

$$(150\text{-}100) / (500/100) = +10$$

that is, 10 more deaths than predicted per 100 patients treated.

The overall predicted number of events at each hospital p is calculated by summing the individual predicted probabilities of an event p_i generated for each patient by the prognostic model ($p = \Sigma p_i$). The method for calculating p_i is shown in Appendix 7 and that for calculating the standard error and 95% confidence interval for a W score are shown in appendix 8.

In order to directly compare hospital outcomes before and after adjusting for casemix, the unadjusted outcome must also be expressed as a W score. An

unadjusted W score is calculated by defining the predicted number of events p at each hospital as np_o, where n is the number of patients at the hospital and p_o is the proportion of patients observed to have experienced the outcome in the SOP overall. (Parry *et al.* 1998b)

The comparison of W scores between hospitals

In the SOP, I was chiefly interested in the comparison of W scores *between* the hospitals rather than the value of the W score at each hospital in isolation. The validity of a comparison of W scores depends on the uniformity of the calibration of the prognostic model, that is, the accuracy of the model predictions for groups of patients across the entire range of predicted risks e.g. low risk, medium risk, high risk groups, etc. This is usually assessed graphically (Harrell, Jr. *et al.* 1996a; Justice *et al.* 1999). In the SOP, I derived calibration plots by stratifying the data set into deciles according to *predicted* risk (0 to 9.9%, 10 to 19.9%, ...90 to 100%) and then plotting the mean predicted risk against the mean observed risk for each decile. For a well calibrated model, a calibration plot should follow a 45° line.

If a model is properly calibrated, model predictions for groups of patients are free from bias and the absolute difference in adjusted outcome between hospitals can be calculated simply by subtracting W scores. For example, if the W score at Hospital X is +3 and that at Hospital Y is –3, then one may infer that there are 6 more events per 100 patients treated at Hospital X than at Hospital Y. However, if the calibration plot deviates substantially from the 45° line, then simple comparisons of W scores

are not valid. To illustrate, consider a prognostic model predicting death that systematically under-estimates risk in groups of patients at high actual risk of death. The greater the proportion of patients at high risk of death admitted, the greater the under-estimate in the overall number of predicted deaths. Any difference of W scores between hospitals will therefore reflect differences in casemix, in this instance, favouring hospitals admitting a smaller proportion of patients at high risk of death. I drew calibration plots for each model to guard against this bias. If the plot largely followed the 45° line, then I calculated the absolute difference in outcome between hospitals simply by subtracting W scores; if the plot deviated markedly from the 45° line, I used the strategy described below.

The Standardised W score

The Standardised W score (Ws score) is a modification of the W score that reduces the bias that results when comparisons in adjusted outcome are made between populations with differing casemix structures when a model used to adjust for casemix fails to calibrate uniformly (Glance *et al.* 2000; Hollis *et al.* 1995; Younge *et al.* 1997). The modification involves standardising the W scores with respect to baseline predicted risk against a reference population; that is to say, the method artificially assigns a similar casemix structure to all hospitals. To do so, the data set at each hospital is first divided into strata according to baseline predicted risk (0 to 9.9%, 10 to 19.9%, etc.) and a W score, termed w_j, calculated for each stratum using the formula:

$$w_j = (o_j - p_j) / (n_j / 100)$$

where o_j is the observed number of events in stratum j, p_j is the predicted number of events in stratum j (calculated as $\Sigma\, p_i$ within stratum j) and n_j the number of cases in stratum j. The standardised W score (Ws score) is then calculated as the product of w_j and f_j (the fraction of the *reference* population in interval j) summed over all strata:

$$w_s = \Sigma_{over\ strata}(w_j . f_j)$$

For relevance, I have used the entire SOP data set as the reference population. Thus, in the SOP, the Ws score is the W score that would result if the casemix structure of each hospital were the same as that of the SOP overall. Since all hospitals are assigned a similar casemix structure, the extent to which their adjusted outcomes are in error because of non-uniform model calibration is also similar, and hence a *comparison* of adjusted outcomes *between* hospitals is valid.

The methods used to calculate the standard error and 95% confidence limits for Ws scores are shown in appendix 8.

3.10 Other statistical methods

I have shown proportions as percentages and, where appropriate, given 95% confidence intervals (calculated using Confidence Interval Analysis, version 1.0) and tested for the significance of any differences using the Chi square test or Fisher's exact test. I have presented odds ratios and relative risk ratios with 95% confidence intervals (calculated using either SPSS for Windows, version 9.0 or Epi Info, version

6.04b). For continuous variables, histogram plots showed that most data were not normally distributed. In the main, therefore, I have presented these data using median and percentile values and used non-parametric methods (either the Mann-Whitney or Kruskall-Wallis tests) to test for the significance of any differences between cohorts (using SPSS for Windows, version 9.0). For data that were normally distributed, I have presented mean values with 95% confidence intervals and tested for the significance of any differences using the independent t-test (SPSS for Windows, version 9.0). Any other calculations were performed using an Excel 97 spreadsheet. All figures were produced either from this spreadsheet or directly from SPSS, version 9.0.

3.11 Data quality

All data were entered into the study database by double punching and verification. I investigated the accuracy of the database prior to analysis using recognised checking procedures (Altman 1999a). Anomalous and missing entries were cross checked against the original audit form and corrected. As a final check on the accuracy of the database, we randomly selected the records of 300 acute stroke hospital stays (~10% sample) and checked *all* entries abstracted from the medical record against the original audit forms. Only seven omissions or mis-punched items were found out of an estimated 20 000 items of data (two describing stroke pathological type, one describing severity and four describing the process of care).

Table 3.1 Characteristics of the five participating hospitals at the start of the SOP

Trust	Characteristics				
	Setting	**Hospital type**	**A&E Dept.**	**Neuro Dept.**	**Stroke services**
A	Urban & rural	DGH	Yes	No	Conventional
B	Urban	Teaching	Yes	No	Organised
C	Rural	DGH	Yes	No	Conventional
D	Urban	Teaching	No	Yes	Organised
E	Urban	DGH	Yes	No	Conventional

DGH District General Hospital
Neuro Neurology and neurosurgery

Table 3.2 The composition of the study hospitals and the referral patterns of stroke patients within each

Hospital	Stroke patients referred from	Reason for referral
Hospital A		
Main hospital	GP and A&E Dept.	Acute care & some rehabilitation
*Subsidiary 1	Main hospital	Acute care and rehabilitation (geriatrics)
*Subsidiary 2	Main hospital	Rehabilitation & long stay beds
*Subsidiary 3	Main hospital	Rehabilitation & long stay beds
*Subsidiary 4	Main hospital	Temporary admission while awaiting a place at a rehabilitation facility; also some long stay beds
Hospital B		
Main hospital	GP and A&E Dept.	Acute care & some rehabilitation
Subsidiary	Main hospital only	Rehabilitation
Hospital C		
Single hospital	GP and A&E Dept.	Acute care & rehabilitation; also some long stay beds
Hospital D		
Main hospital	GP only Surrounding hospitals	Acute care & rehabilitation Neurology & neurosurgery
Subsidiary	GP & Main hospital	Acute care & rehabilitation (geriatrics)
Hospital E		
Single hospital	GP and A&E Dept.	Acute care & rehabilitation; also some long stay bedsϕ

* These hospitals officially belonged to another NHS Trust. However, patients admitted to Hospital A with stroke are routinely transferred to these hospitals for rehabilitation and sometimes acute care. Patients that were directly admitted to these hospitals with stroke are *not* included in this study.

ϕ The long stay wards at Hospital E belong to another NHS Trust.

Table 3.3 ICD Cerebrovascular Disease Codes, ninth and tenth revisions

ICD 9 codes	**Definition**
430*	Subarachnoid haemorrhage
431*	Intracerebral haemorrhage
432.0/1	Non-traumatic extradural/subdural haemorrhage
432.9	Unspecified non-traumatic intracranial haemorrhage
433*	Occlusion and stenosis of precerebral arteries
434*	Occlusion of cerebral arteries
435*	Transient cerebral ischaemia
436*	Acute but ill-defined cerebrovascular disease
437*	Other and ill-defined cerebrovascular disease
438*	Late effects of cerebrovascular disease
ICD 10 codes	
I60*	Subarachnoid haemorrhage
I61*	Intracerebral haemorrhage
I62.0/1	Non-traumatic extradural/subdural haemorrhage
I62.9	Unspecified non-traumatic intracranial haemorrhage
I63*	Cerebral infarction
I64*	Stroke, not specified as haemorrhage or infarction
I65 / I66*	Occlusion and stenosis of precerebral/cerebral arteries, *not* resulting in cerebral infarction
I67*	Other cerebrovascular diseases
I68*	Cerebrovascular disorders in diseases classified elsewhere
I69*	Sequelae of cerebrovascular disease
G45*	Transient cerebral ischaemia

* all suffixes

Table 3.4 The information routinely reported for stroke patients on the SMR1 (Scottish Morbidity Record Type 1)

About the patient:

Hospital case reference number

Surname in full; forenames *initials only*

Date of birth

Sex

Marital status

Post code *only* of their home address *prior* to admission

About their illness:

Principal diagnosis (the main condition managed or investigated during the patient's stay)

Up to five subsidiary diagnoses (complications and co-morbid conditions)

About their General Practioner:

A unique code number identifying the address of the GP practice

The GP's General Medical Council registration number *(optional)*

About their consultant:

A unique code identifying the consultant responsible for the patient whilst in hospital

The medical specialty of the consultant.

About their admission:

Where they were admitted from e.g. home, other hospital etc.

Type of admission e.g. emergency, transfer, waiting list, etc.

Dates of admission and discharge

Discharge code e.g. to home, to other hospital, died in hospital, etc.

Type of hospital facility e.g. in-patient, day case, etc.

Table 3.5 Information abstracted from the medical record and entered onto the study database in all cases regardless of diagnosis

Patient identification

Date of birth

Sex

Postcode of usual place of residence (used to derive socio-economic status)

Marital state (by SMR1 definitions †)

About the admission

Date of first admission

Date of first clerking in the medical record

Diagnostic category (stroke, ineligible stroke, not stroke, not known)

Diagnosis if not a stroke

† SMR1 defintions of marital status:

Never married/single
Married, includes sperated
Widowed
Other, includes divorced
Not known

Table 3.6 Definition of variables used to describe casemix in patients with an audited diagnosis of acute stroke

Prior to the stroke	
Diabetes mellitus	Past history of diabetes mellitus
Ischaemic heart disease	Past history of angina, myocardial infarction or coronary artery bypass graft surgery
Myocardial infarction	Past history of a myocardial infarction
Lived alone	Patient regularly lived alone
Independent in ADLs	Able to perform everyday activities without assistance e.g. walking, bathing, feeding, dressing
Employed	In regular part or full time employment
Within first 24 hours of admission	
Orientated & able to speak	Speech understandable and orientated in time and place (i.e. normal verbal component of the GCS)
Normal GCS Eye score	Eyes open spontaneously
Normal GCS Motor score	Able to move unaffected/ least affected limb purposefully
Able to lift both arms	MRC motor score grade 3 or more in both arms
Able to walk without help	Able to walk without human assistance i.e. the use of walking aids was acceptable
High systolic blood pressure	Systolic blood pressure over 160 mmHg
Urinary incontinence *	Two or more episodes of urinary incontinence or any use of an indwelling catheter *

* within the first seven days of admission

Table 3.7 Data collected by the survey of the structure of stroke care

Organisation & specialisation of care

Geographically defined stroke unit

If present: Type of unit: acute and/or rehabilitation

On/ off main hospital site

Number of beds/ ability to expand

Date opened

Physician specifically responsible for patients with stroke

If yes: membership of stroke related interest groups

Provision and use of a stroke specific clerking proforma

Wards & staff usually involved in providing stroke care

Number of beds per ward

Nurses: number and grade per ward

Therapists: number and grade assigned to medical wards/stroke unit

Stroke liaison nurse

Social services

Access to diagnostic facilities

CT and MRI head scan

Carotid artery doppler ultrasound scan

Echocardiography

Video-fluoroscopy

Outpatient

Access to a day hospital and therapists

Provision of specialist neurovascular clinic for patients with TIA and minor stroke

Other

Written protocol for the treatment of stroke patients within the hospital

Maintenance of a hospital stroke register

Table 3.8 Data describing the process of care collected in cases of acute stroke

Principal measures (from the medical record)

Organisation of care

- multidisciplinary team (MDT) meeting held; date of first
- admitted to a geographically defined stroke unit; date first admitted
- discharged from the care of a physician with a special interest in stroke

Discharge on an antithrombotic medication (aspirin or anticoagulant)

CT Head scan; date of first scan

Supplementary measures (from the medical record)

Use of subcutaneous heparin; date started

Use of parenteral fluids; date started

Insertion of a urinary catheter; date inserted

Stroke units only (by hospital survey/ auditors experience only)

Age related admission policy

Provision of education and training for staff

Provision of education and support for patients and carers

Setting and documenting of rehabilitation goals

Formal assessment/ re-assessment of progress of rehabilitation

Frequency, attendance and documentation of MDT meetings

Discharge policy

Table 3.9 The predictive variables of the models used to predict outcome after stroke

1. Age on admission

Prior to the stroke

2. Was the patient independent in activities of daily living?
3. Did the patient live alone?

On admission

4. Was the patient orientated and able to speak?
5. Was the patient able to lift both arms?
6. Was the patient able to walk without help?

Table 3.10 Details of the logistic regression models used to predict outcome at six months after stroke.

Model covariate	Coefficient (se)
1. Survival at six months	
Constant	9.2647 (1.3918)
Age (per year)	-0.0428 (0.0148)
Lived alone	0.6921 (0.2957)
Independent	-0.7511 (0.3446)
Orientated and able to speak	-1.3886 (0.3070)
Able to lift both arms	-1.1551 (0.3640)
Able to walk without help	-0.9155 (0.3707)
2. Alive and independent at six months	
Constant	12.3397 (1.4818)
Age (per year)	-0.0512 (0.0130)
Lived alone	0.6612 (0.2606)
Independent	-2.7442 (0.5144)
Orientated and able to speak	-2.1603 (0.4727)
Able to lift both arms	-2.1062 (0.4732)
Able to walk without help	-1.3109 (0.3774)
3. Alive and at home at six months	
Constant	9.0428 (1.2579)
Age (per year)	-0.0500 (0.0132)
Lived alone	1.1819 (0.2675)
Independent	-1.1667 (0.3443)
Orientated and able to speak	-1.1746 (0.3319)
Able to lift both arms	-1.4041 (0.3619)
Able to walk without help	-1.3459 (0.3327)

Coding of variables

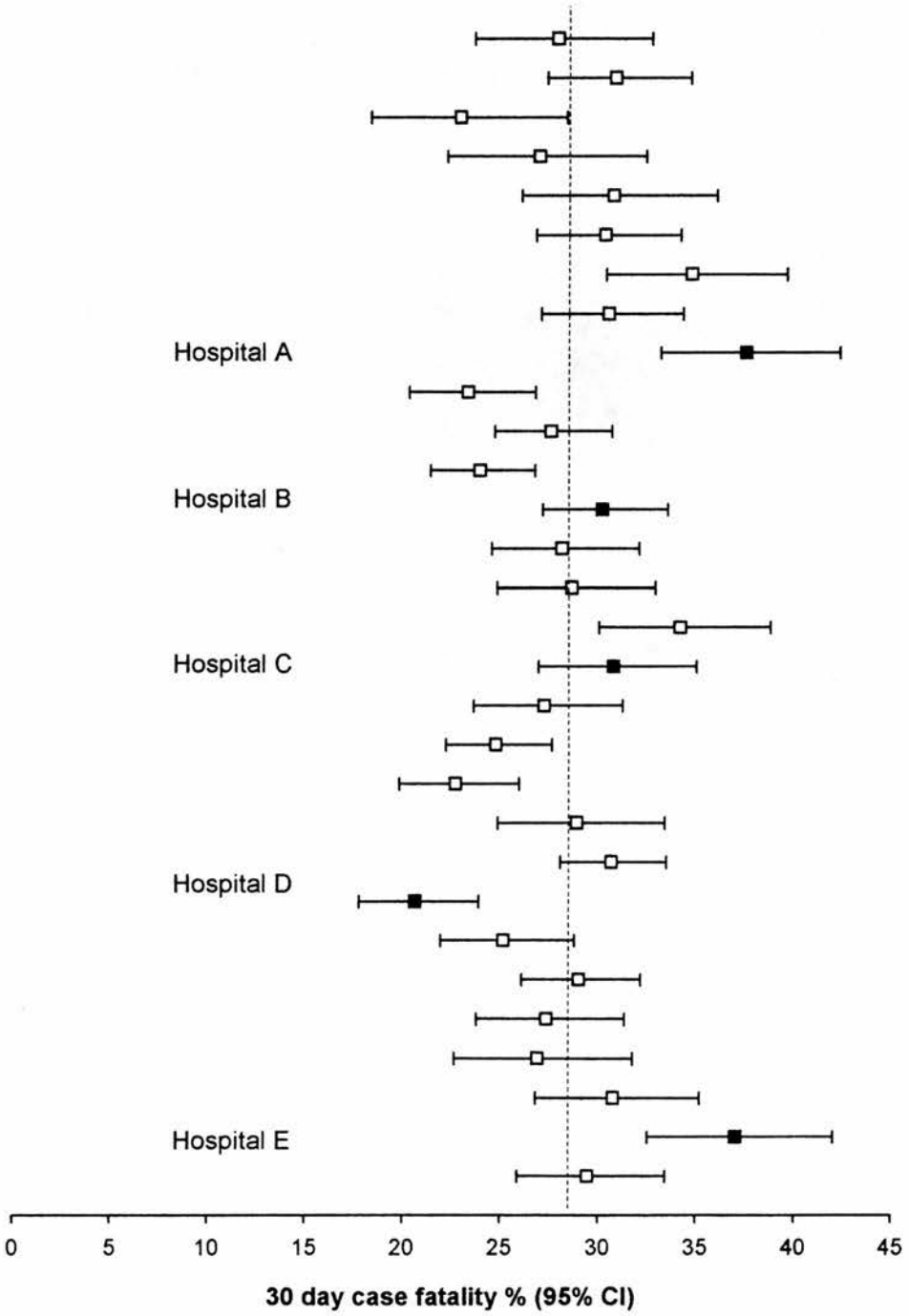
Outcomes: alive, 'alive and independent', 'alive and at home' = 1;
dead, 'dead or dependent', 'dead or not at home' = 0.
i.e. the models predict the probability of good outcomes.

Covariates: presence of the attribute = 1, absence of the attribute = 2.

The predicted probability of death or of 'death or dependency'
= (1 - the predicted probability of being alive or 'alive and independent', respectively).

Figure 3.1 Deaths within 30 days of emergency admission for stroke for each Scottish hospital for the period July 1990 to June 1993 as published by the Clinical Outcomes Working Group of the Scottish Office in December 1994 (corrected in February 1995)

The proportion dead (with 95% condifence interval) for each hospital is shown by a point and two whiskers; the dotted line represents the mean proportion dead in Scotland overall (28.4%)



Chapter Four. How accurately do routine hospital discharge data identify patients with acute stroke?

4.1 introduction

Hospital discharge data currently provide the only practical means of identifying patients with a given disorder routinely and across an entire country. These data are easily available and are generated unobtrusively and automatically, nearly all being derived from the clinical summaries that are sent to the GP shortly after a patient's discharge. Unfortunately, these practical advantages are offset by the well recognised potential for inaccuracy (Demlo and Campbell 1981b; Harley and Jones 1996; McKee and James 1997b; Williams and Mann 2002). First, only some of the patients labelled as having the disorder in question may in fact do so (i.e. there may be false positive error) and second, the data may identify only a proportion of the total number of the patients admitted with the disorder (i.e. there may be false negative error). Both errors reduce the credibility of the data and, more importantly, may lead to bias and reduced power in any comparison of outcome between hospitals. If the intention is to follow patients to determine their functional status after discharge, false positive error also increases the number of patients to be followed.

This chapter describes an attempt to determine the accuracy with which routine discharge data identified patients with acute stroke. This investigation is primarily related to false positive error. The identification of error requires a gold standard

against which the study data can be compared: to identify false positive error, I have used the diagnosis data collected by our research assistant as the final arbiter of whether or not a patient had an acute stroke; to identify false negative error (sensitivity), we would have had to identify *all* patients with stroke at each hospital independently of their discharge data i.e. we would have had to set up five prospective stroke registers. This was beyond our resources. Nonetheless, false negative error resulting from the selection of *subsets* of CVD codes can be studied and the potential impact of any reduced sensitivity in CVD discharge data as a whole can be estimated.

I have investigated false positive error in CVD discharge data per hospital and per ICD code. For the per hospital analyses, I have concentrated on the accuracy of the subsets of CVD codes used by ISD and by the Department of Health (DoH) to derive the stroke outcomes data currently published in Scotland and in England & Wales, respectively, and attempted to quantify the impact of such error on comparisons of case fatality between hospitals. I have also summarised the reasons for admission in patients with a CVD code as the primary diagnosis but who did not have an acute stroke and briefly considered various other aspects of SMR1 data that pertain to case identification and which might impact on the ability to follow patients at six months.

4.2 Methods

CVD codes and SMR1 data

All analyses are based on SMR1s that had a CVD code in the *primary* diagnostic position (see Table 3.3). When the same hospital stay was reported using SMR1s with differing CVD codes I categorised the stay by the CVD code given on the *final* SMR1. I categorised hospital stays as emergency or non-emergency admissions according to the type of admission code given on the *first* SMR1. We collected *all* CVD discharge diagnosis codes only for the first six months of the study (during which time the majority of SMR1s used ICD9 codes). All analyses refer to data collected over this period except for that describing the accuracy of the individual ICD codes which is based upon the entire period the codes were collected (which ranged from six months to two years). ISD collects all emergency admissions with any CVD code except TIA listed as the primary SMR1 diagnosis whilst the Department of Health collects all emergency admissions with ICD 10 codes I61 to I64 listed as the primary diagnosis (analogous to ICD 9 codes 431, 432, 434 & 436).

False positive and negative error

I quantified the false positive error in terms of positive predictive value (PPV), calculated as the proportion of all hospital stays that were verified as having been for acute stroke. I quantified false negative error in terms of the proportion of hospital stays for an acute stroke that were lost from analysis.

Acute stroke

All analyses in this chapter refer only to hospital stays for acute stroke, ineligible stroke and not stroke. For the purpose of calculating PPV, I have classified patients with ineligible stroke (see section 3.5.1 for definition) as having a non-stroke disorder. Thus, the estimates of PPV indicate the proportion of hospital stays for patients who presented within four weeks of stroke onset and who were not treated elsewhere first i.e. patients by whom a hospital's 'performance' with regard to the treatment of acute stroke care might fairly be judged.

4.3 Results

In the first six months, we inspected the medical record and identified a diagnosis in 97% (1255) of the 1291 hospital stays with a CVD code listed as the primary SMR1 diagnosis: 713 acute strokes; 64 ineligible strokes; 478 not strokes. Using the ISD method of identification there were 638 acute strokes, 24 ineligible strokes and 177 not strokes; using the DoH method there were 578 acute strokes, 11 ineligible strokes and 118 not strokes.

4.3.1 Positive predictive value per hospital

The results of the per hospital analyses are shown in Table 4.1. When all CVD codes were used to identify acute stroke, PPV was poor and varied significantly between hospitals (range 51 to 69%, p 0.0007). When the ISD method of case identification was used, PPV improved (overall PPV 76%; range 70 to 82%, p 0.146). The DoH

method of identification was somewhat better (overall PPV 82%) and the variation between hospitals smaller (78 to 85%, p 0.596). The ISD method of identification resulted in a loss of 11% of audit verified acute strokes (range 2% to 17%). The DoH method of identification resulted in an overall loss of 19% of audit verified acute strokes (range 10% to 32%).

4.3.2 The positive predictive value of individual CVD codes

The PPV of individual CVD codes are shown in Table 4.2. The great majority of hospital stays relating to acute stroke were assigned one of four codes: intra-cerebral haemorrhage (431/I61); 'ischaemic stroke' (434/I63); acute unspecified stroke (436/I64); and non-specific cerebrovascular disease (437/I67). The PPV of the first three codes were relatively high (range 80 to 87%) but that for 437/I67, the third most commonly used CVD code, was only 26%.

The code for SAH (430/I60) had a moderate PPV for acute stroke (69%). However, of the 125 hospital stays for an acute stroke that were assigned this code, 86% had in fact had a SAH i.e. the code had a very low PPV for *non-SAH* acute stroke.

The code for non-traumatic intracranial haemorrhage (432/I62) appeared to have only a moderate PPV for acute stroke (50%). However, the PPV of the sub-code referring to non-specific intracranial haemorrhage (432.9/I62.9) was high (93%;14/15 cases). The low PPV for the code overall related to the inclusion of sub-

codes for non-traumatic subdural/extradural haemorrhage (I62.0/I62.1) which had a PPV for acute stroke of only 7% (1/15 cases).

Three CVD codes had very low PPV for acute stroke: transient ischaemic attack (code 435; PPV 5%); stenosis or occlusion of precerebral arteries (code 433; PPV 1%) and late effects of cerebrovascular disease (code 438; PPV 0%). In order to increase the efficiency of our audit of the medical records, we stopped collecting these codes after the first six months of the study. Similarly, we stopped collecting hospital stays with a SAH code listed as the primary SMR1 diagnosis after the first year.

The overall PPV of the 980 hospital stays coded as an emergency admission was 66% while that for the 275 hospital stays coded as a non-emergency admission was 24%.

4.3.3 Modifying the ISD method of identifying acute stroke

With the knowledge that certain CVD codes had a low PPV for acute stroke, I explored the possibility of improving the PPV of the ISD method by omitting cases with the ICD9 codes 430, 433 & 438 (a method which equates to the DoH method plus code 437). The benefit of this modification was minimal (overall PPV 77%) and was essentially limited to Hospital D. Here, the PPV of its cohort improved

from 73 to 79% but also resulted in the loss of 19% of the acute strokes identified using the standard ISD method.

4.3.4 Impact of the measured coding errors on comparisons of case fatality

When any CVD code was used to identify patients with acute stroke, the *absolute* estimates of six month case fatality were uniformly and substantially low (refer forward to Table 7.19, page 276). The *absolute* estimates of six month case fatality improved at all hospitals when the standard ISD and the DoH methods of patient identification were used. The estimates of case fatality at Hospitals B & D were essentially correct using each method and those at Hospitals A,C & E were moderate under-estimates (up to 4% in absolute terms) except at Hospital A where the ISD method led to an absolute under-estimate of 7%. The modified ISD method gave very similar estimates of case fatality as the standard ISD method.

The *comparisons* of case fatality between hospitals were in reasonable agreement with the truth except for the comparisons between Hospitals A & B (for both the ISD and DoH methods) and between Hospitals A & D (for the ISD method only). Thus, the ISD method suggested that case fatality at Hospital B was 12% lower than at Hospital A and the DoH method suggested that the difference in case fatality was 17%. In fact, the gold standard suggested that case fatality at Hospital B was 26% lower than at Hospital A. Similarly, the ISD method suggested that case fatality at Hospital D was 40% lower than that at Hospital A when in fact it was 47% lower.

4.3.5 Actual diagnosis in non-stroke cases assigned a CVD code

The actual reason for admission in 478 patients assigned a CVD code as their primary SMR1 diagnosis but who did not have an acute or ineligible stroke is shown in Table 4.3. The reason for admission fell into the following categories: other cerebrovascular disease (almost exclusively TIA) (29%); an investigation or procedure (nearly all relating to a prior stroke or TIA) (16%); a specific neurological symptom or condition (25%); a non specific and/or non neurological condition (25%); and unknown disorder/missed stroke (5%). The notes made by our audit assistant showed that the specific neurological symptoms or conditions and the non-specific and/or non neurological conditions often either presented in a manner that resembled stroke or occurred on a background of established cerebrovascular disease. The ISD method reduced the identification of non-stroke patients across all categories but especially those with TIA and those who had been admitted for an investigation or procedure.

4.3.6 Other findings

Less than one percent of the post codes referring to the patient's pre-admission address differed from that extracted from the medical record. These few errors were in the last three symbols of the code, not the area codes. 77% (550) hospital stays consisted of only one finished consultant episode (FCE), 17% (119) consisted of two FCEs and only 6% (44) consisted of three or more FCEs. The median delay between admission to hospital and ISD providing us with the SMR1 was 137 days; we were

informed of 95% of relevant SMR1s by 177 days (i.e. within six months.) and 98% by 270 days (i.e. within nine months).

4.4 Discussion

These analyses suggest that if all SMR1s with any CVD code listed as the primary diagnosis are used to identify acute stroke, the resulting hospital samples are very impure with between one half and one third of hospital stays referring to another disorder. This finding is not really a surprise. It is, after all, the purpose of CVD codes to describe cerebrovascular diseases other than acute stroke. It was therefore to be expected that our samples would also contain patients with TIA, SAH, less common cerebrovascular pathologies and, given the coding convention of diagnosis first and procedure second, some with carotid artery stenosis admitted for angiography, angioplasty or endarterectomy. Also, mis-coding of non-cerebrovascular disorders using CVD codes is well recognised. As in other studies, these disorders generally presented in a manner that resembled or might commonly be misdiagnosed as stroke or occurred on the background of established cerebrovascular disease (Davenport *et al.* 1996c; Goldstein 1998; Mant *et al.* 1997; Stegmayr and Asplund 1992). The prognosis of all these conditions is generally better than that of acute stroke and hence it was also not a surprise that the estimates of case fatality based on these hospital cohorts were uniformly and substantially low.

The ISD and the DoH methods of selecting subsets of CVD codes are designed to increase the proportion of acute strokes in each hospital cohort and to improve the

accuracy of their estimates of outcome. Our findings confirm the rationale of each method in that the excluded codes all had a low PPV for acute stroke. As a result of their exclusion, the PPV of the hospital cohorts improved substantially, more so with the DoH method than with the ISD method. The lower PPV of the ISD method primarily relates to its inclusion of code 437/I67, which, although often used to code acute stroke, appears to be even more often used to code for a 'rag-bag' of non-stroke disorders. These findings are in keeping with other studies that have investigated the PPV of subsets of CVD codes (Goldstein 1998; Harley and Jones 1996; Liu *et al.* 1999; Mant *et al.* 1997).

However, neither the DoH nor the ISD methods were especially powerful and about a fifth to a quarter of cases identified by them still referred to patients without an acute stroke. The practical consequence of this is that any system of routine follow up based upon them would be inefficient and would run the risk of alienating patients, public and GPs who might not understand why they, their relatives or patients were being needlessly questioned. From a measurement perspective, the PPV of the hospital cohorts continued to vary appreciably using the ISD method of patient identification and both, but especially the DoH method, led to non-uniform and in some cases large losses of patients with acute stroke. Despite this, most estimates and comparisons of case fatality were only minimally in error. However, at some hospitals, and especially using the ISD method, the degree of error was more substantial. In particular, the ISD method underestimated the relative difference in case fatality between Hospitals A & B by 14%. This degree of error is equivalent to about two-thirds of the treatment effect that might be expected if one hospital were to

admit *all* its patients to a stroke unit and the other *none at all* (Langhorne *et al.* 1993a), which is to say a degree of error that might mis-lead a study that aimed to measure quality of stroke care using case fatality data. The slightly more accurate estimates and comparisons of case fatality obtained using the DoH method of identification suggest that it should be used in preference to the ISD method, although, by also giving smaller sample sizes and hence wider confidence intervals, it may be less able to differentiate between hospitals with truly different outcomes.

An important caveat to all these conclusions is that this study is small, both in terms of the number of hospitals studied and, because we collected all CVD codes only for six months, in terms of the numbers of patients in each cohort. As a result, it is possible that my estimates of the degree of bias in measurements of case fatality and of the number of hospitals where this bias is large may themselves be influenced by the play of chance. Hence, my findings should be viewed cautiously and, ideally, a larger study should be performed to confirm or refute them. Also, my findings only describe accuracy in relation to the population of strokes that had a CVD code assigned as the primary SMR1 diagnosis i.e. they ignore the other major potential source of bias, namely false negative coding error.

The impact of false negative coding error can be estimated from the literature and by inference. Mant (1997) summarised a number of single hospital studies which investigated the accuracy of routine data systems for acute stroke in the UK and found that their sensitivity varied from as low as 66% to as high as 97% (although

some of this apparent variation may have related to differences in the number of diagnoses and subsets of CVD codes that were collected). Acute strokes appeared to be 'missed' by routine systems for a number of reasons, including (Barer *et al.* 1996; Davenport *et al.* 1996c; Leibson *et al.* 1994; Mant *et al.* 1997; Panayiotou *et al.* 1993; Stegmayr and Asplund 1992):

- coding the stroke as a symptom or sign e.g. hemiplegia;
- coding the stroke as 'unknown cause of morbidity/mortality' because a clinical diagnosis was unavailable to the coding staff;
- coding the stroke as a secondary diagnosis with a complication or risk factor given as the primary diagnosis;
- coding the stroke as a secondary diagnosis because it occurred during an admission for another disorder;
- coding a risk factor or complication whilst failing to code the stroke at all;
- misdiagnosing and hence mis-coding minor stroke as TIA (where TIA codes are excluded from the subset under collection);
- and in cases of rapidly fatal stroke, those who die in the accident & emergency department and hence have not been formally admitted to the hospital.

In some cases, therefore, false negative coding may relate to stroke severity and/or treatment failure and hence to outcome. On this basis, variation between hospitals in the sensitivity of their routine CVD discharge data may indeed bias the comparison of outcomes. A *rough approximation* of the potential impact of this bias can be

made by taking, as an example, a hospital with 1000 admissions for acute stroke, a true observed and predicted case fatality of 30% (i.e. a ratio of observed (O) to predicted (P) deaths of 1.0) and, amongst the false negative cases, a clinically plausible range of baseline severity and quality of care (see appendix 9 for methods). The findings are shown in Table 4.4.

At 90% sensitivity, the error in overall case fatality appears to be minimal. At 80% sensitivity, the error is only appreciable when the unreported strokes are substantially more severe and/or receive substantially lower quality of care than the reported cases. However, at 70% sensitivity, the error is appreciable even when unreported strokes are only moderately more severe and/or receive only moderately lower quality of care than reported cases. On balance, it seems that most hospitals in the UK (and elsewhere) produce routine CVD discharge data with a sensitivity of greater than 80% (Mant *et al.* 1997) and therefore, in most cases, the impact of reduced sensitivity is unlikely to be large. However, should sensitivity be closer to 70%, as it appears may sometimes be the case (Mant *et al.* 1997), then estimates and comparisons of outcome after stroke may be appreciably in error.

The bias resulting from reduced sensitivity may be partly overcome by adjusting comparisons of outcome for casemix. Here, the predictive model provides a standard (the predicted outcome) against which the outcome of each patient may be compared (e.g. observed/predicted) to produce an index of performance that is independent of the proportion of patients whose outcome is reported. This ability to 'correct' for

reduced sensitivity appears to hold true regardless of how insensitive the CVD discharge data are and regardless of how different the prognosis of the unreported cohort is from that of the reported cohort (Table 4.4). However, it clearly does *not* hold true if the quality of care provided to the unreported cohort is different to that of the reported cohort (Table 4.4). Furthermore, the process of adjusting for casemix may itself be a source of bias given the inevitable presence of false positive coding error and the fact that the predictive models developed for patients with stroke are unlikely to be applicable to those with non-stroke disorders.

Thus, no simple statement can be made about the accuracy with which routine CVD discharge data identify patients with acute stroke. Using the DoH or ISD methods, the bias due to false positive coding error seems likely to be modest at many hospitals but at some, perhaps the minority, it may be sufficiently large to mislead studies which aim to measure quality of care. The potential for false negative coding bias is less certain but it appears that it may be significant at some hospitals and it is difficult to tell how far adjustment for casemix can compensate for it. A further problem, thus far un-remarked, is that the coding of stroke is unrefined. The great majority of acute strokes are assigned the code 436/I64.9 (acute strokes not defined as a haemorrhage or infarct) preventing the proper identification of their pathological sub-type, a problem well recognised by others (Ellekjaer *et al.* 1999; Leibson *et al.* 1994; Liu *et al.* 1999; Mant *et al.* 1997) and which may hamper attempts to use outcomes data to explore the quality of care in more detail e.g. what proportion of patients with ischaemic stroke were discharged on an antithrombotic drug. No amount of shuffling of the remaining CVD codes can overcome any of these

difficulties which can only be addressed, and hence truly confident comparisons of outcome between hospitals only be made, if the quality of routine coding of acute stroke is improved.

Several factors contribute to the current poor quality of routine CVD data. By far the most important is the poor quality of the information supplied by physicians to coding clerks, usually on a hand-written discharge letter or a typed discharge summary (Davenport *et al.* 1996c; Harley and Jones 1996; Hasan *et al.* 1995; Mant *et al.* 1997; Panayiotou *et al.* 1993; Patel *et al.* 1976a). The job of providing these data frequently falls to junior doctors almost none of whom have received any formal training, who often have no idea that their statements form the basis of important official statistics, who view the task as a chore that often has to be completed outside of working hours and who get little or no feedback on the quality of their efforts (Frain *et al.* 1996). As a result, it is not a surprise that the required information may sometimes fail to materialise or is greatly delayed, contains mis-diagnoses or mis-ordered diagnoses, or is incomplete, vague or illegible.

Other less important factors also play their part in the poor quality of routine discharge diagnosis data. Although trained, lack of medical knowledge may lead coding staff to mis-interpret the diagnosis or the order of diagnoses (if they are forced to do so) (Panayiotou *et al.* 1993; Yao *et al.* 1999) and lack of time may lead coding staff to use less refined codes or to make simple errors; subtle differences in coding policy may also exist between hospitals (Harley and Jones 1996). The ICD

CVD codes themselves are also problematic (McKee and James 1997a), in particular 437/I67 (other/ill-defined cerebrovascular disease) some of whose sub-categoreis have pathological definitions that have no clear clinical counterpart and which can also be 'unspecified' (I67.9) providing a pigeon-hole for vague diagnostic terms.

Thus there are a number of ways by which the quality of routine discharge diagnostic data might be improved. Coding staff clearly require training, support, up-dates and quality control both within and between departments, items which in fact are already in place in Scotland (Harley and Jones 1996). Clinically based coding systems, such as Read codes, may make the translation of the physicians terms into a pathologically based ICD code easier, although Read codes themselves are not necessarily straightforward to use and their implementation is patchy. However, the key to improving the quality of discharge data would be to improve the quality of the data provided by physicians. Unfortunately, issuing guidelines on best practice (Anonymous 1990) and exhorting physicians to do better is unlikely to bring this about. Rather, it will require their education and commitment and also the commitment of hospital management (Davenport *et al.* 1996c; Frain *et al.* 1996; Panayiotou *et al.* 1993; Patel *et al.* 1976b; Williams and Mann 2002; Yao *et al.* 1999).

Physicians should be educated, perhaps starting at medical school, about routine data systems and their importance, and trained in approved methods of recording clinical information both in the medical record and in discharge documentation (in particular,

the importance of assigning a clear primary diagnosis and the use of symptoms/signs rather than vague terms when the primary diagnosis is not clear). Physicians should also be prepared to discuss cases, resolve difficulties with coding staff and respond to feedback on the quality of their work. More accurate diagnostic coding would also result if physicians became involved in their assignment themselves (Yeoh and Davies 1993), especially if senior doctors were involved (Leibson *et al.* 1994). This extra commitment would clearly be time consuming. It should therefore be matched by the provision of protected time within the working day and, where relevant, electronic look-up systems should be provided. Unfortunately, none of these suggestions are new, indeed some are over 25 years old (Patel *et al.* 1976b). The failure to instigate them probably reflects the real world of the NHS where there are many more immediate priorities, the failure (until recently) of analyses based upon routine data to impact on clinicians, and the fact that the implementation of these recommendations would require culture change on the part of physicians (Williams and Mann 2002; Wyatt 1995). Maybe now, with a the imminent introduction of new electronic data systems within the NHS, some progress may be made.

Finally, it is necessary to consider the other aspects of SMR1 that are important in identifying cases and which allow a routine system of follow up to contact survivors at six months after admission. As noted, currently the only address data on the SMR1 is the post code of the patients address prior to admission and, fortunately, it appears this is very accurate. Although not directly studied, other demographic data and admission and discharge dates on the SMR1 are also likely to be correct in 97 to 99% of cases (Harley and Jones 1996; Kohli and Knill-Jones 1992). SMR1s are

clearly returned by hospitals sufficiently quickly that a follow up questionnaire can be mailed before six months in the great majority of cases and in virtually all by nine months (when measured in large groups, functional outcome changes little between six and nine months after stroke (Skilbeck *et al.* 1983)). Lastly, about a quarter admissions are reported using more than one SMR1, confirming the need to link SMR1s to prevent the despatch of multiple follow ups. ISD report a high degree of accuracy in their ability to link SMR1s to make hospital stays (Kendrick and Clarke 1993) but their linkage is periodic and has never been tested in an on-going basis. Nonetheless, our experience suggests that it is possible. An alternative solution might be to respond only to the first SMR1 received on any patient and to ignore any others received later. However, arrangements would have to be made to deal with SMR1s arriving in the wrong order and to halt follow up when subsequent SMR1s indicate that the patient is dead. In the long run, better still might be to abandon the concept of consultant episodes and to return to a simpler and more logical system of simply reporting finished episodes of care, which for stroke, would be a single hospital stay (Clarke and McKee 1992; Williams and Mann 2002).

Summary

1. Only about half to two-thirds of patients in Scotland with any CVD code listed as the primary diagnosis have had an acute stroke.

2. The positive predictive value for acute stroke of the methods of identification used by ISD and the DoH is 75 to 80%. The use of either method leads to a substantial loss of patients with acute stroke.

3. Most estimates of case fatality derived using the ISD and DoH methods of case identification are only modestly biased (in terms of false positive coding error). However, in some cases, and especially using the ISD method, the error may be large enough to mislead efforts to measure the quality of care using outcomes data. These conclusions require confirmation by a larger study.

4. This study had not addressed the question of variation in sensitivity (false negative coding error) of routine CVD discharge data. The literature suggests that sensitivity is usually greater than 80%, a level where bias seems likely to be low. If the sensitivity falls to 70% then only moderate differences between reported and unreported cases in stroke severity or quality of care may lead to important bias in estimates and comparisons of outcome.

5. Adjustment for casemix reduces bias due to reduced sensitivity *provided* the unreported and reported cases have received care of the same quality. However, by the very nature of false negative reporting in stroke, it is dangerous to assume that this is the case.

6. Truly confident comparison of outcome after stroke requires improvement in the accuracy of CVD discharge data. This is only likely to occur if clinicians improve the quality of clinical data reported in routine discharge data and/or take part in assigning codes themselves.

Table 4.1 The positive predictive value for acute stroke of the hospital cohorts derived from CVD discharge data

	Positive predictive value of SMR1 CVD data for acute stroke												
	1. All CVD codes		2. ISD method		3. DoH method		4. ISD method (modified)		Acute strokes 'lost'				
									1.→ 2.		1.→ 3.		1.→ 4.
	stroke/ all cases	% (95% CI)	stroke/ all cases	% (95% CI)	stroke/ all cases	% (95% CI)	stroke/ all cases	% (95% CI)	n	%	n	%	n %
Hospital A	122/177	**69** (62-76)	120/153	**78** (72-85)	109/129	**85** (78-91)	116/147	**79** (72-86)	2	**2**	13	**11**	6 **5**
Hospital B	214/402	**53** (48-58)	178/234	**76** (71-82)	171/214	**80** (75-85)	173/229	**76** (70-81)	36	**17**	43	**20**	41 **19**
Hospital C	144/233	**62** (56-68)	130/159	**82** (76-88)	121/143	**85** (79-91)	122/148	**82** (76-89)	14	**10**	23	**16**	22 **15**
Hospital D	144/283	**51** (45-57)	122/167	**73** (66-80)	98/120	**82** (75-89)	99/126	**79** (71-86)	22	**15**	46	**32**	45 **31**
Hospital E	89/160	**56** (48-63)	88/126	**70** (62-78)	79/101	**78** (70-86)	85/120	**71** (63-79)	1	**1**	10	**10**	4 **4**
TOTAL	713/1255	**57** (54-60)	638/839	**76** (73-79)	578/707	**82** (79-85)	595/770	**77** (74-80)	75	**11**	135	**19**	118 **17**

ISD Information & Statistics Dept. DoH Department of Health

ISD method: SMR1s with any CVD codes except TIA admitted as an emergency

DoH method: SMR1s with CVD codes I61 to I64 (analogous ICD9 codes 431,432,434 & 436) admitted as an emergency

ISD method (modified): SMR1s with CVD codes I61 to 64 & I67 (analogous ICD9 codes 431,432,434,436 & 437) admitted as an emergency

Table 4.2 The positive predictive value of the individual ICD cerebrovascular disease codes for acute stroke

Cerebrovascular disease code ICD 9 and 10 codes combined			**Audited diagnosis** n				**Positive predictive value** % (95% CI)
	ICD9	ICD10	Acute stroke	Ineligible stroke	Not stroke	Total	
'Ischaemic Stroke'	434	I63	775	36	79	890	**87** (85-89)
Intracerebral Haemorrhage	431	I61	267	27	39	333	**80** (76-85)
Acute Unspecified Stroke	436	I64	1543	61	310	1914	**81** (79-82)
SAH †	430	I60	125	27	30	182	**69** (62-75)
Non-traumatic Intracranial Haemorrhage	432.9	I62	15	3	12	30	**50** (31-69)
Non-specific CVD	437	I67	111	6	312	429	**26** (22-30)
TIA†‡	435		8	0	147	155	**5** (2-10)
Stenosis or Occlusion of Precerebral Artery ‡	433		1	0	79	80	**1** (0-7)
Late effects of CVD ‡	438		0	0	4	4	**0** (0-60)
Total (n 4017)			2845	160	1012	4017	

'Ischaemic stroke' : code 434 = occlusion of intracerebral arteries; code I63 = cerebral infarction

† SAH codes collected only for one year
‡ Codes collected only for six months

Table 4.3 The diagnoses of cases assigned a CVD code as the primary SMR1 diagnosis but which were not found to have had an acute stroke.

Audited diagnosis	All cases		Emergency, non-TIA cases only	
	%	n	%	n
1. Other cerebrovascular diagnoses	**29**	138	**15**	27
TIA		136		27
Other		2		0
2. Investigation or procedure	**25**	120	**2**	3
Angiography / endovascular intervention/other investigation for stroke, TIA or SAH		79		1
Pre-operative assessment		4		0
CEA/other vascular procedure		17		0
PEG tube insertion/management		7		2
Other		13		0
3. Neurological symptom or disorder	**16**	84	**30**	54
Headache (including 'exclude SAH')		7		7
Cerebral tumour		4		3
Head injury		5		4
Vertigo		3		0
Dementia		11		7
Seizures		16		13
Syncope/dizziness/'VBI'		12		3
Collapse/loss of consciousness		7		4
Miscellaneous		19		13
4. Non-specific or non-neurological disorder (intercurrent or non-specific illness, confusion, poor mobility, inability to cope, falls, fractures, etc.)	**25**	118	**49**	86
5. Probably was acute or ineligible stroke *	**1**	3	**1**	2
6. Unclassifiable	**1**	3	**1**	2
7. No details recorded	**3**	12	**2**	3
TOTAL		**478**		**177**

VBI Vertebro-basilar artery ischaemia
CEA Carotid endarterectomy
PEG Percutaneous endoscopic gastrotomy

* In these cases our research assistant recorded *erroneously* classified venous sinus thrombosis and stroke admissions over 4 weeks after symptom onset as a non-stroke disorder. The former are acute strokes and the latter are 'ineligible' strokes.

Table 4.4 An illustration of the impact of reduced sensitivity in the reporting of admissions for acute stroke

Example assumes 1000 admissions for acute stroke, a true case fatality of 30% and a true adjusted outcome in line with predictions (O/P =1)

Quality of care in excluded strokes * (ratio of O/P deaths; in reported cases O/P = 1.00)	Severity of excluded strokes (% predicted to die)	Sensitivity of discharge data (%)	Strokes identified	Deaths		Crude case fatality (%)	Adjusted case fatality (O/P)
				Observed	Predicted		
		100	1000	300	300	**30**	**1.00**
Same as reported cases							
1.00	30	90	900	270	270	**30**	**1.00**
		80	800	240	240	**30**	**1.00**
		70	700	210	210	**30**	**1.00**
1.00	40	90	900	260	260	**29**	**1.00**
		80	800	220	220	**28**	**1.00**
		70	700	180	180	**26**	**1.00**
1.00	50	90	900	250	250	**28**	**1.00**
		80	800	200	200	**25**	**1.00**
		70	700	150	150	**21**	**1.00**
Worse than reported cases							
1.10	30	90	900	267	270	**30**	**0.99**
		80	800	234	240	**29**	**0.98**
		70	700	201	210	**29**	**0.96**
1.10	40	90	900	256	260	**28**	**0.98**
		80	800	212	220	**27**	**0.96**
		70	700	168	180	**24**	**0.93**
1.10	50	90	900	245	250	**27**	**0.98**
		80	800	190	200	**24**	**0.95**
		70	700	135	150	**19**	**0.90**

Worse still than reported cases							
1.20	30	90	900	264	270	**29**	**0.98**
		80	800	228	240	**29**	**0.95**
		70	700	192	210	**27**	**0.91**
1.20	40	90	900	252	260	**28**	**0.97**
		80	800	204	220	**26**	**0.93**
		70	700	156	180	**22**	**0.87**
1.20	50	90	900	240	250	**27**	**0.96**
		80	800	180	200	**23**	**0.90**
		70	700	120	150	**17**	**0.80**

* A worse quality of care results in more deaths than predicted (according to baseline characteristics) and hence O/P ratios greater than one. An O/P ratio of 1.1 implies 10% more deaths than predicted, etc.

Chapter Five. How reliably can the variables of our predictive models be collected?

5.1 Introduction

Accurate adjustment for casemix using a statistical model can only be made if the predictive variables included in the model are valid (measure what they purport to measure) and reliable (consistent on repeated measurement) (Laupacis *et al.* 1997c; Wyatt and Altman 1995c). Should detailed adjustment of stroke outcomes ever become routine, the necessary predictive variables will have to be collected by a clinician at the time of the patient's admission (prospectively) or by an auditor inspecting the information that was laid down in the medical record (retrospectively). The predictive variables included in our prognostic models (see Table 3.9) are, by design, simple and purely clinical and so ought to lend themselves to routine collection. The aim of this chapter is to determine whether or not this is the case by estimating the reliability of our predictive variables when prospectively collected and their reliability and validity when retrospectively collected. Urinary incontinence and conscious level are other simple and important predictors of outcome after stroke (Gladman *et al.* 1992a; Henon *et al.* 1995; Irwin and Rudd 1998; Ween *et al.* 1996) which might also be used to adjust stroke outcomes for casemix on a routine basis; indeed, both are currently included in the Royal College of Physicians minimum data set for stroke (Irwin and Rudd 1998). Urinary incontinence is particularly important, being able to predict death, disability and residence after stroke more accurately than conscious level alone (Wade and Hewer 1985) and at least as well as some published

multivariate prognostic models (Barer and Mitchell 1989; Gladman *et al.* 1992b; Taub *et al.* 1994). A further aim of this chapter, therefore, is to assess the reliability and validity of data describing urinary incontinence and conscious level (the latter in terms of GCS eye score, normal vs. abnormal, a rough approximation for normal or reduced level of consciousness).

Reliability can be assessed either by comparing the values assigned on more than one occasion by the same observer (intra-rater reliability) or by comparing the values assigned by two or more different observers (inter-rater reliability). I chose to measure inter-rater reliability because, in a routine setting, our prognostic data would necessarily be collected by many different observers. This chapter therefore describes two straightforward inter-rater studies, one prospective (clinician vs. clinician) and the other retrospective (auditor vs. auditor). To determine the validity of the retrospectively collected predictive data, it was necessary to compare them against a gold standard. I defined the gold standard to be the same data collected prospectively by an independent and expert observer, close to the time the retrospective data were collected. Thus, I determined the validity of our retrospectively collected data by means of a third inter-rater study (clinician vs. abstractor). Similar methods have been used to validate the retrospective collection of the variables of the Canadian Neurological Scale (Bushnell *et al.* 2001; Goldstein and Chilukuri 1997) and the National Institutes of Health Stroke Scale (Baird *et al.* 2000; Bushnell *et al.* 2001; Kasner *et al.* 1999; Williams *et al.* 2000).

There were two main reasons to doubt the validity of our retrospectively collected predictive data. First, from an audit and research perspective, the quality of the data ordinarily laid down in the medical record may be poor. Thus, the recording of baseline characteristics may be incomplete, inaccurate, and lacking in detail; the entries may be contradictory, illegible or in code; and the records themselves may be disordered or partly or wholly missing. As a result, the collection of predictive data often involves a process of decipherment and inference and hence the potential for error. Second, inspection of the medical record often identifies the hospital to which the patient was admitted and/or the patient's outcome and this knowledge may consciously or unconsciously bias the collection of data (Caplan *et al.* 1991a). It is of course possible that data collected prospectively may also be invalid since clinicians may make incorrect assessments. However, in view of the simplicity of the predictive variables studied, I suspected that any such bias would be small. Moreover, the potential for systematic disagreement between prospective observers (i.e. biased measurement by one) can be estimated from a study of reliability (see below).

5.2 Methods

5.2.1 Study 1. Prospective inter-rater reliability (clinician vs. clinician)

This study was limited to Hospital D. Here, two neurology registrars with an interest in stroke (NUW and CEC) assessed a consecutive series of patients admitted with an acute stroke between March and September 1997. We identified patients by daily contact with the admitting physicians and by monitoring cases referred for CT head

scan. We tried to assess the patient as soon as possible after each other and always did so on the same day, which, in nearly all cases, was the day after admission. We gathered data by interviewing and examining patients directly; where this was not possible (e.g. due to dysphasia or reduced conscious level) we collected the missing variables from the medical record or by interviewing nursing staff and relatives. All data were collected using a standardised form and blind to the findings of each other.

5.2.2 Study 2. Retrospective inter-rater reliability (auditor vs. auditor)

This study was performed at all five hospitals in the SOP. Two observers (NUW and the study research assistant, AG) assessed the medical records, including the nursing entries, of a sample of patients eligible for the study. We derived our sample by selecting every n^{th} patient on the SOP database with n varying per hospital in order to provide 60 patients at each. Each observer first determined whether or not the patient had been diagnosed to have an acute stroke (see section 3.6.1) and then collected the predictive variables, plus data describing stroke pathological sub-type (see section 3.6.2). This study therefore also allowed me to estimate how reliably we identified cases of acute stroke and their pathological sub-type in the SOP. All data were collected blind to the findings of the other auditor and using forms with the same wording and prompts.

5.2.3 Study 3. Prospective vs. retrospective reliability (retrospective validity)

This study was also limited to Hospital D. Here, I extracted prospectively collected predictive data from the hospital stroke register, these data having been originally collected by one of five research registrars with a special interest in stroke using the same methods as described under Study 1 (see section 5.2.1). The retrospectively collected data were extracted from the medical record, including nursing entries, by our research assistant and pertained to the day of admission. Ideally, both prospective and retrospective examinations should relate to the same day. However, the majority of patients on the stroke register were not examined till the day after admission or even later still. As a compromise, therefore, I restricted this analysis to patients in whom prospective data were collected within 48 hours of admission.

5.2.4 Definitions and data entry

For prospective assessments, we defined acute stroke by the WHO criteria (Asplund *et al.* 1988b) and included cases within 30 days of stroke onset. For the retrospective assessments, we defined acute stroke using the SOP criteria (see section 3.5.1). The definitions of the predictive variables have already been given (see Table 3.6). All data were entered onto the study database by means of double punching and verification except, because of limited resources, the data collected by NUW in the retrospective reliability study (see section 5.2.2). These data were entered using only single punching (but the database was thoroughly checked for anomalous values).

5.2.5 Measurement of agreement

Continuous data

The only continuous prognostic variable that we collected was age. I used the method proposed by Altman to quantify the inter-observer agreement (Altman 1999b). To do so , I first calculated the difference in the age assigned by each observer to the same patient and then calculated the median value of such differences over all patients, the 'median difference'. The median difference along with the 5th to 95th percentiles summarise the range of disagreement between the two observers.

Categorical data

For categorical data, I have expressed inter-rater reliability using the proportion of agreement between observers and using the kappa statistic (κ) which expresses the degree of agreement achieved *beyond* that attributable to chance. The concept of agreement beyond that attributable to chance is best understood graphically (see Figure 5.1). Kappa is calculated using the formula

$$\kappa = (p_o - p_e) / (1 - p_e)$$

where p_o is the observed proportion of agreement and p_e is the proportion of agreement attributable to chance (Altman 1999b). The standard error (and hence confidence interval) for kappa can be derived using a standard formula (Fleiss *et al.* 1969). Kappa is constrained to vary from –1, which indicates perfect disagreement, to +1, which indicates perfect agreement. By convention, κ values between 0 and 0.20 indicate slight agreement, between 0.21 and 0.40 fair agreement, 0.41 and 0.60

moderate agreement, 0.61 and 0.80 good agreement, and 0.81 to 1.00 excellent agreement (Altman 1999b; Brennan and Silman 1992).

Whilst this hierarchy provides a useful rule of thumb, it is important to realise that very different contingency tables can result in similar κ values; κ values should therefore always be interpreted in the context of the raw data from which they were derived (Brennan and Silman 1992). Two points should be borne in mind. First, kappa values are dependent on the prevalence of the attribute in question: a high prevalence necessarily results in high chance agreement. Here, high observed levels of agreement may be accompanied by surprisingly low kappa values. The dependence of kappa on the prevalence of the attribute also means that simple comparisons between populations where the prevalence varies may not be valid. Second, the observed level of agreement reflects both random variation and any systematic differences between observers i.e. chance and bias. As a summary statistic, kappa is unable to separately identify the influence of bias. Bias may, however, be identified by the presence of imbalance in the 'off diagonal' cells in the contingency table. The significance of any suspected bias can be estimated using McNemar's test (Brennan and Silman 1992). This test computes a z statistic of the null hypothesis that there is no bias; the p value of the null hypothesis may then be derived from standard statistical tables.

5.3 Results

5.3.1 *Study 1. Prospective inter-rater reliability (clinician vs. clinician)*

Both observers examined 93 patients at a median of two days after stroke onset (inter-quartile range 1 to 4 days; range 1 to 17 days). The mean time between examination by NUW and by CEC was 3.5 hours (range 0.3 to 7 hours). One patient had a further stroke in the interval between examinations and hence has been excluded from analysis. Also, we did not collect data on urinary incontinence in the first six patients. Therefore, all analyses refer to our findings in 92 patients except for urinary incontinence which is based on our findings in 86. The median predicted risk of death by six months (by NUW ratings) in the study cohort was 15% (IQR: 4 to 40%) i.e. the majority had mild to moderate stroke.

The median difference in the assessment of age was zero years (5^{th} to 95^{th} percentiles: 0 to 0 years) (see Table 5.1). The minor disagreements were the result of small differences in our assessment of the day of stroke onset; the two larger discrepancies (1.18 and 1.99 years) were the result of difference in our assessment of date of birth. In both these cases, the patients were confused. We collected the categorical variables of our study models with good to excellent reliability (Table 5.2). Agreement was lowest on pre-stroke independence in the activities of daily living (ADLs) (κ 0.67). Some of the disagreement on this variable was systematic (z 2.41, p 0.016); NUW was less likely to find that the patient had been independent than CEC. Agreement on urinary incontinence and on GCS eye score was excellent.

5.3.2 *Study 2. Retrospective inter-rater reliability (auditor vs. auditor)*

The identification of cases for this study is shown in Figure 5.2. Of the 298 cases originally identified, both auditors inspected the medical records of 274 (92%). Of these, nine had been included in error (they were admitted directly to one of the affiliated units at Hospital A, see section 3.3) and so were excluded from analysis. The reliability of the diagnosis of stroke is therefore based on a sample of 265 cases; the reliability of the collection of predictive variables and of the diagnosis of stroke pathological subtype is based on the 200 cases where both NUW and AG agreed on the diagnosis of acute stroke. The median delay between stroke onset and the first recording of predictive data in the medical record was zero days (inter-quartile range 0 to 1 days; range 0 to 20 days); ten patients had a stroke whilst already admitted for another condition. The median predicted risk of death by six months (by NUW ratings) in the cohort of 200 cases was 48% (IQR: 16 to 77%) i.e. the study included patients with a broad range of stroke severity.

The median difference in the assessment of age was zero years (5^{th} to 95^{th} percentiles: 0 to 0 years). Disagreement was generally only a matter of one to two years but in one instance it was marked (10 years) (see Table 5.1). Discrepancies were due to transcription error by the observers and to differences in dates in different parts of the medical record. We collected the categorical variables of the study models with moderate to excellent reliability (Table 5.2) although the level of agreement was less than that achieved by the two prospective observers (except for the variable 'living alone'). Agreement was lowest on the ability to walk without

assistance (κ 0.55) and second lowest on independence in ADLs (κ 0.61). Disagreement on independence in ADLs and on orientation and ability to speak was systematic (z 2.01, p 0.044 and z 2.89, p 0.004, respectively); in both cases, NUW was less likely to find the patient to have been independent or orientated and able to speak than AG. Agreement on urinary incontinence was excellent and that on the GCS eye score was good . Agreement on the diagnosis of acute stroke was good (κ 0.76) and almost perfect on the diagnosis of stroke pathological sub-type (κ 0.95) (Table 5.3).

5.3.3 Study 3. Prospective vs. retrospective reliability (retrospective validity)

We collected predictive data both retrospectively (relating to the day of admission) and prospectively (within 48 hours of admission) in 195 patients (see Figure 5.3); only 35 patients had both sets of data collected on the day of admission. Excluding six cases without a date of stroke onset and six cases where stroke occurred whilst already in hospital, the median delay from stroke onset to admission was zero days (IQR: 0 to 1 day, range 0 to 21 days). The median predicted risk of death by six months (by NUW ratings) in the cohort of 195 cases was 11% (IQR: 5 to 43%) i.e. the majority had mild to moderate stroke.

The median difference in the assessment of age between retrospective and prospective observers was zero years (5^{th} to 95^{th} percentiles: 0 to 0 years). Any disagreements on age were minor (see Table 5.1). Agreement on the categorical variables included in our models ranged from moderate to good (Table 5.2).

Agreement was least for independence in ADLs (κ 0.49) with disagreement being partly systematic (z 4.23, $p < 0.0001$); the prospective observer was less likely to find that a patient had been independent than the retrospective observer. Agreement on urinary incontinence and on GCS eye score was good. Some of the disagreement on urinary incontinence was systematic (z 4.63, $p < 0.0001$); the retrospective observer was more likely to judge the patient to be incontinent than the prospective observer.

5.4 Discussion

This study shows that the six variables included in our prognostic models and data describing urinary incontinence and conscious level (in terms of the normality or otherwise of the GCS eye score) can be collected very reliably by clinicians at the patient's bedside. It also suggests that the same data remain reliable and reasonably valid when retrospectively collected from the medical record. Thus, the variables included in our prognostic models, urinary incontinence and GCS eye score (normal vs. otherwise) may indeed be suitable for routine collection, whether by prospective or retrospective means, and hence to the routine adjustment of stroke outcomes for casemix. The study also shows that we were able to identify acute stroke and its pathological sub-types very reliably in the SOP as a whole.

The satisfactory reliability of the variables included in our prognostic models is likely to reflect the decision to exclude, as far as was possible, variables with known or presumed low reliability (e.g. sensory impairments) and variables which are

informative in only a small proportion of patients (e.g. bilateral extensor plantar reflexes) during model development (Counsell *et al.* 2002). We achieved the lowest level of inter-rater agreement over the three assessments for the variable describing pre-stroke independence in ADL (κ 0.49 - 0.67). In each assessment, disagreement between observers was partly systematic. Discussion revealed that this was because of variation between observers in the definitions of ADL and dependency. More reliable assessments might have been possible if we had used a checklist to specify which ADL to consider and defined the threshold at which the patient should be considered dependent. While ADL are often taken to include washing, dressing, feeding, toileting and mobilising (Wade 1995), it is less clear, for instance, whether bathing or shopping should be included since these are not necessarily daily activities. Indeed, a definition of functional independence which excludes bathing and shopping would probably be sensible given the importance of environmental factors in determining abilities in these areas (i.e. bath or shower; distance from shops). Similarly, there should be clear agreement between observers on how to assess dependency in the presence of alterations to the patient's environment (Burn 1992).

As noted, comparisons of inter-rater data derived from different populations should be performed cautiously. Nonetheless, as one might expect, we found generally better agreement between observers when data were collected prospectively than when they were collected retrospectively. In particular, the ability to walk unaided was extremely reliable when assessed prospectively but only moderately reliable when assessed retrospectively. Reviewing the medical records showed that this

discrepancy was probably due to the infrequency with which physicians specifically record the ability to walk soon after admission. Often, the ability to walk unaided had to be inferred from other information that was recorded by the physicians (e.g. leg strength) and/or by the nursing staff. Although the nursing records usually contained information about mobility, it was frequently in terms of the need for supervision and thus was difficult to interpret. For example, it was difficult to tell whether the patient needed 'one to help with walking' because they could not walk alone or because the nurse felt they were potentially unsafe if left to their own devices; moreover, whether the nurse provided physical support or simply walked alongside was often unclear.

These observations support the contention that if clinical predictive data are ever to be routinely collected, it would be preferable for clinicians to explicitly record them in the medical record at the time of admission using standard set of definitions. This would be greatly facilitated if the predictive variables were included on a stroke clerking proforma (Davenport *et al.* 1995a). Not only would such prospective collection be more accurate than retrospective collection, it would also be much cheaper (see section 1.5.2). Certainly, a pilot study that investigated the ability of coding clerks (in Scotland) to collect our six casemix variables showed that it would require a great deal of training and supervision to avoid error (Dr Jill Peel, personal communication). However collected, a very important secondary benefit of the routine collection of clinical predictive data is that it would improve the accuracy with which routine hospital discharge data are able to identify cases of acute stroke: those cases with predictive data appended would, by definition, be highly likely to

have been admitted with an acute stroke and therefore to receive an appropriate diagnostic code. As such, the creation of a system to collect predictive data would reduce the need to educate and persuade clinicians to provide higher quality clinical information in discharge summaries, a task that has proved rather thankless over many years (see section 4.4)

Routine transfer of predictive data from the hospital to the centre would require that hospital discharge data be capable of reporting descriptive clinical variables. Although not currently possible in England and Wales, Scotland is fortunate that those responsible for designing the up-dated version of the SMR1 (introduced in the mid 1990s) had the foresight to include the capacity to report up to six condition specific casemix or outcome variables (Anonymous 1995), which is to say, a system that would allow routine transfer of our predictive variables already exists. Even so, this system would rely on a chain of doctors and coding clerks and would be both inefficient and open to error. In the longer term, the introduction of an electronic patient record (Burns 1998) has the potential to transform the collection of predictive data by forcing clinicians to enter the variables on the day of admission and then by automatically appending the data to the SMR1.

Given their importance as predictors of outcome after stroke, it is reassuring that data describing urinary incontinence and conscious level are reliable and reasonably valid when collected retrospectively. Indeed, our data probably under-estimate the validity of retrospectively collected incontinence data: our prospective data relate only to the

first two days of admission while the retrospective data relate to the first seven days, hence the systematic tendency for the retrospective observer to find incontinence when the prospective observer did not. Previous studies suggest that the reliability of the assessment of level of consciousness varies from fair to excellent (D'Olhaberriague *et al.* 1996; Dewey *et al.* 1999). Higher levels of agreement appear to result when observers have been given special training. Inexperienced observers have little difficulty identifying patients at the extremes of consciousness (i.e. normal or very low) but are unreliable and inaccurate in the middle range (Rowley and Fielding 1991a). The Royal College of Physicians stroke minimum data set (Irwin and Rudd 1998) currently proposes the worst level of consciousness in the first 24 hours as a marker of stroke severity, measured using a loosely defined four point scale (fully conscious, drowsy, semi-conscious, unconscious). Given that these data would likely be recorded by junior physicians without special training, it is arguable that the use of a simple and reliable dichotomous variable, such as whether or not the GCS eye score was normal, might be preferable.

The principal shortcoming of all three studies is the extent to which their findings can be confidently generalised to other settings. Doubts in this regard refer to the observers, hospitals, types of patient and timing of assessment that were used. Thus, the variables were collected either by trainee neurologists with an interest in stroke or by an experienced audit assistant with a nursing background; two studies were performed at a single hospital and, because this hospital lacked an Accident & Emergency department, included very few patients with severe stroke; the same studies also made the majority of their observations on the day after admission.

Hence, it cannot be assumed that less experienced observers (whether junior hospital doctors or coding clerks) dealing with less alert and less co-operative patients during the 'hurley-burley' of admission would make equally accurate assessments (Rowley and Fielding 1991b; Shinar *et al.* 1985). In other words, our findings cannot confidently be generalised to the settings that would apply if the predictive variables were routinely collected for the purposes of routinely adjusting stroke outcomes for casemix. Nonetheless, our findings are an important step in this direction and suggest that accurate data collection might at least be possible. Ideally, the next step would be to repeat the studies described here, heeding the need for more explicit definitions and methods of recording, using a less selected group of patients, clinicians, auditors and hospitals. A useful addition to these studies would be the inclusion of experienced clinicians whose assessments might act as a 'gold standard' against which the accuracy of the assessments of the less experienced observers might be measured (Rowley and Fielding 1991a). That said, given the relative simplicity of our six casemix variables and the encouraging findings of the studies that were performed, it seems unlikely that their collection by junior doctors would be seriously in error.

Summary

1. The variables included in our prognostic models and data describing urinary incontinence and GCS eye score (normal vs. otherwise) appear to be very reliable when prospectively collected and reasonably reliable and valid when retrospectively collected under the conditions of the study. The validity of retrospectively collected urinary incontinence data may be even higher than that shown.

2. It is likely that the reliability of data collection would be improved if our variables were more explicitly defined and if they were clearly recorded in the medical record at the time of admission. For the purpose of routine data collection this might best be achieved if the predictive variables were included in a stroke clerking proforma.

3. Although encouraging, the estimates of the reliability and validity of the predictive variables cannot be confidently generalised to the routine setting. Further study using less specialised observers and a wider range of patients and hospitals is indicated.

4. The routine collection of predictive data would, by default, also improve the accuracy with which routine hospital discharge data are able to identify cases of acute stroke.

Table 5.1 Agreement between observers on age in the three inter-rater studies

	Difference in age between observers (years)	Number of patients
Prospective data study		
clinician vs. clinician	- 1.99	1
	- 1.18	1
	- 0.01	2
	0	80
	0.01	8
		Total 92
Retrospective data study		
auditor vs. auditor	- 10.00	1
	- 2.00	1
	- 1.00	4
	- 0.10	1
	- 0.08	1
	- 0.02	1
	- 0.01	1
	0	181
	0.01	2
	0.08	2
	0.33	1
	1.00	2
	1.04	1
	1.99	1
		Total 200
Validity of retrospective data study		
prospective vs. retrospective collection	-0.01	1
	0	192
	0.01	1
	0.60	1
		Total 195

Table 5.2 Agreement between observers for the categorical variables included in the three inter-rater studies

Predictive variable	Prospective reliability study n 92					Retrospective reliability study n 200					Retrospective validity study n 195				
		NUW		Agreement			NUW		Agreement			Prosp.		Agreement	
	CEC	Yes	No	%	kappa (95% CI)	AG	Yes	No	%	kappa (95% CI)	Retro.	Yes	No	%	kappa (95% CI)
Lived alone	Yes	31	2	92	**0.84** (0.72,0.95)	Yes	68	5	97	**0.92** (0.87,0.98)	Yes	60	4	90	**0.78** (0.68,0.87)
	No	5	54			No	2	125			No	16	115		
Independent in ADLs	Yes	65	10	88	**0.67** (0.50,0.85)	Yes	160	15	90	**0.61** (0.46,0.76)	Yes	151	25	86	**0.49** (0.33,0.65)
	No	1	16			No	5	20			No	2	17		
Orientated & able to speak	Yes	53	4	90	**0.79** (0.66,0.92)	Yes	83	19	*86	**0.73** (0.63,0.82)	Yes	117	17	†83	**0.60** (0.48,0.73)
	No	5	30			No	8	89			No	16	44		
Able to lift both arms	Yes	70	1	99	**0.97** (0.91,1.00)	Yes	90	5	94	**0.88** (0.81,0.95)	Yes	137	6	89	**0.71** (0.59,0.83)
	No	0	21			No	7	98			No	15	37		
Able to walk without help	Yes	48	0	96	**0.91** (0.83,1.00)	Yes	23	13	87	**0.55** (0.40,0.70)	Yes	66	19	81	**0.61** (0.50,0.73)
	No	4	40			No	14	150			No	18	92		
Urinary incontinence	Yes	27	0	‡94	**0.87** (0.76,0.98)	Yes	111	11	93	**0.86** (0.78,0.93)	Yes	50	31	83	**0.62** (0.51,0.73)
	No	5	54			No	3	75			No	3	111		
Eyes open spontaneously	Yes	87	1	99	**0.88** (0.66,1.00)	Yes	132	14	*89	**0.74** (0.64,0.85)	Yes	168	3	94	**0.70** (0.53,0.87)
	No	0	4			No	7	46			No	8	15		

Table 5.3 Agreement between auditors on the diagnosis of acute stroke and its pathological subtype in the retrospective inter-rater study

1. Diagnosis of acute stroke

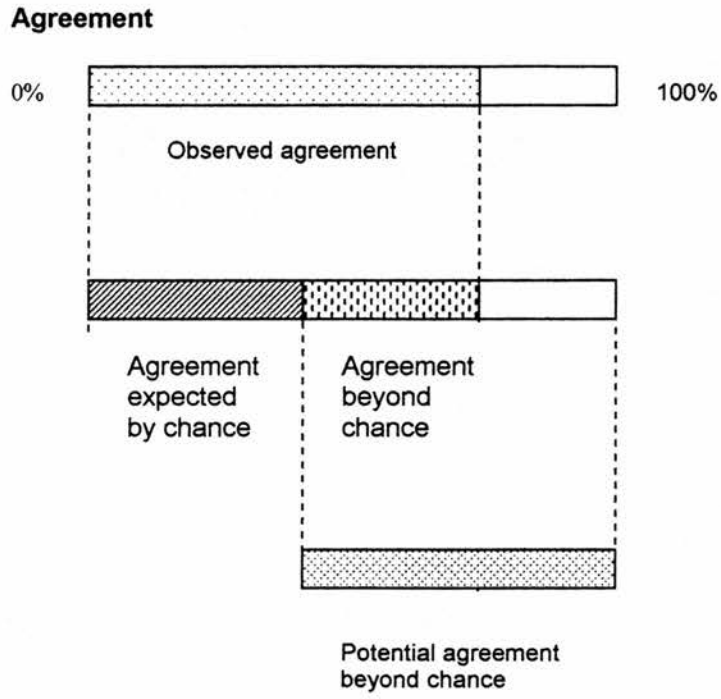
n 265				Agreement	
	NUW			%	kappa (95% CI)
AG	Acute stroke	Not stroke	Ineligible stroke		
Acute stroke	200	5	1	**92**	**0.76** (0.67, 0.85)
Not stroke	7	38	2		
Ineligible stroke	6	1	5		

2. Diagnosis of stroke pathological sub-type

n 200				%	kappa (95% CI)
	NUW				
AG	Ischaemic stroke	ICH	Not known		
Ischaemic stroke	130	1	1	**98**	**0.95** (0.91, 0.99)
ICH	1	24	1		
Not known	1	0	41		

ICH Intracerebral Haemorrhage

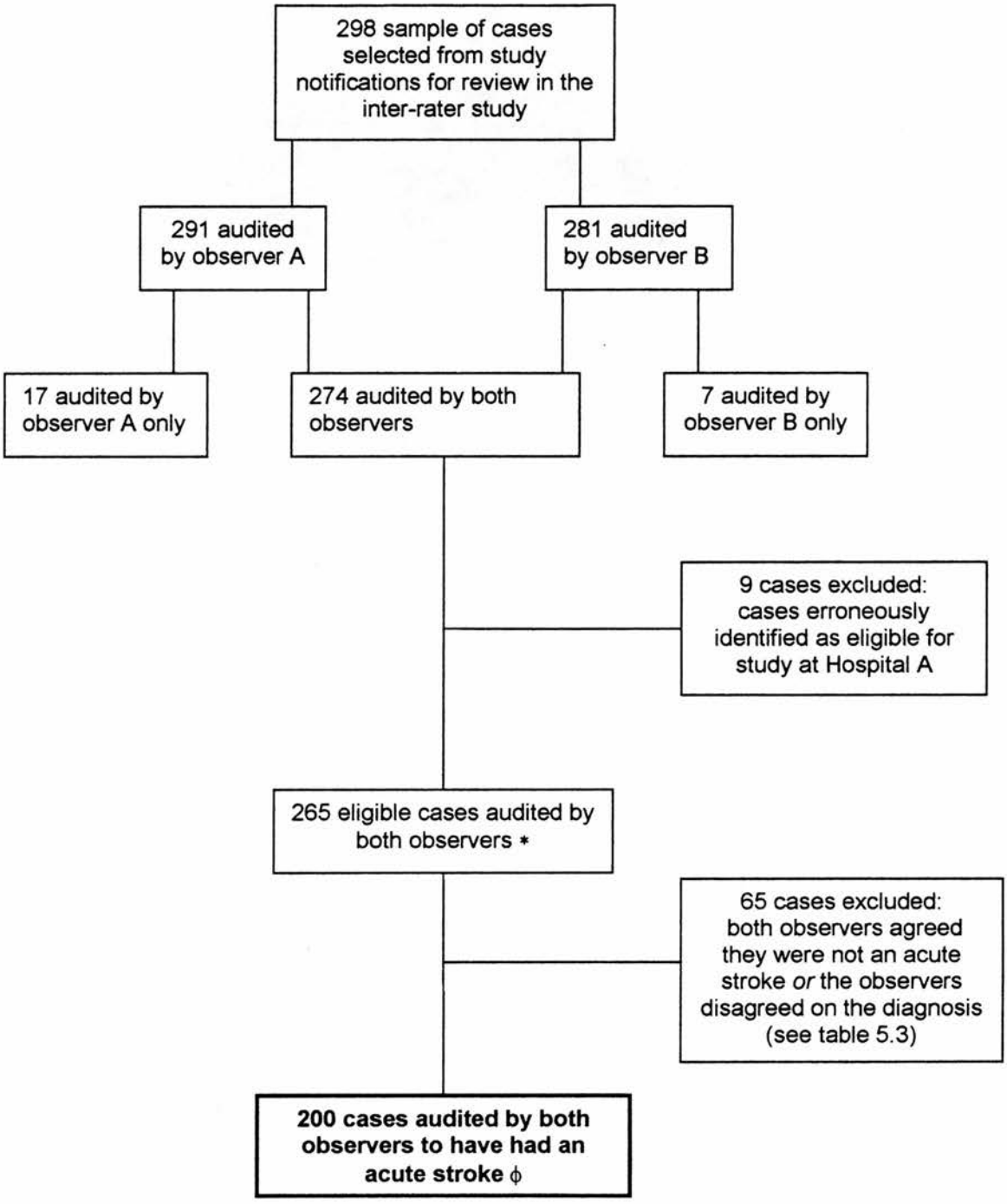
Figure 5.1 A graphical representation of agreement measured by the kappa statistic



$$\text{Kappa} = \frac{\text{actual agreement beyond chance}}{\text{potential agreement beyond chance}}$$

Figure adapted from Clinical Epidemiology: A basic science for medicine, D Sackett, RB Haynes, P Tugwell, Little, Brown and Company, Boston/Toronto, 1985

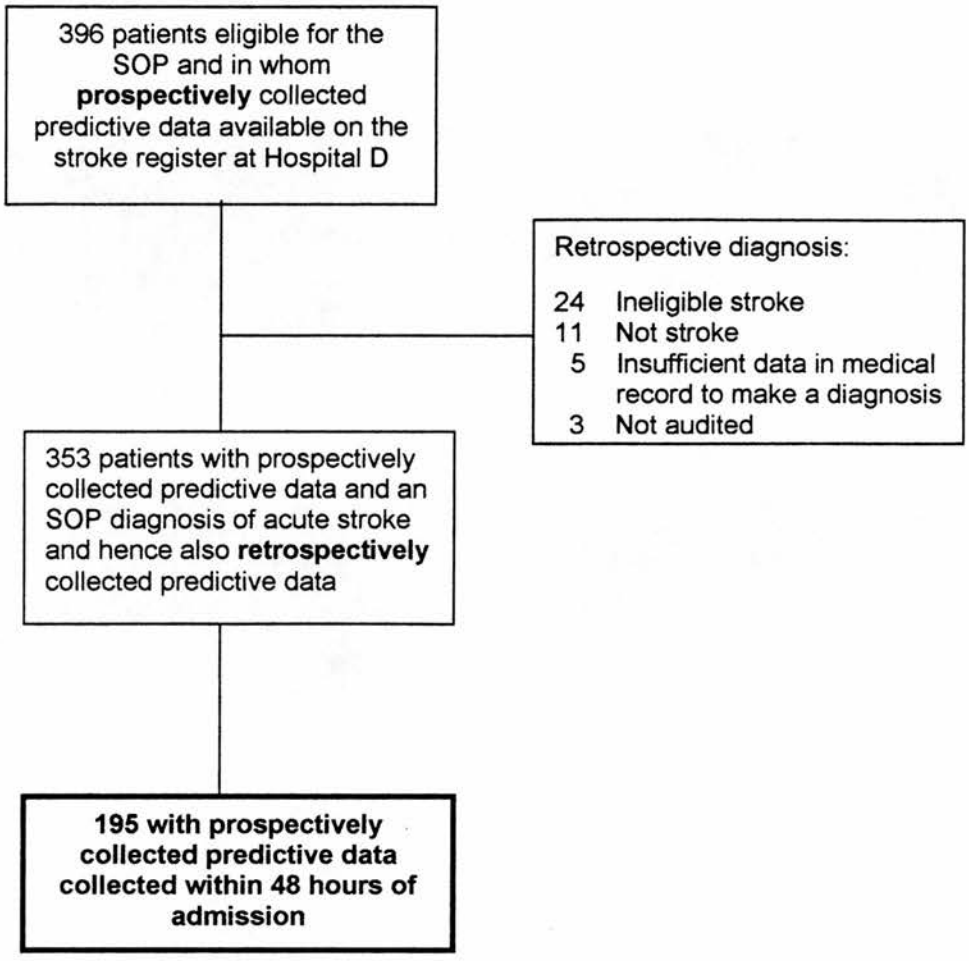
Figure 5.2 Deriving the sample for the retrospective reliability study



* Sample used to estimate reliability of the diagnosis of stroke

ϕ Sample used to estimate reliability of collection of the prognostic variables

Figure 5.3 Deriving the sample for the prospective vs. retrospective reliability study



Chapter 6. Can the response to our outcome questionnaire be improved and what is the impact of non-response?

6.1 Introduction

A high follow up rate is important to any study which aims to measure patient outcomes for two reasons. First, as Sackett *et al* (1985) put it, "..patients do not disappear from a study for trivial reasons", which is to say, patients often fail to respond to follow up for reasons that are associated with the outcome under investigation. Non-response is therefore a potential source of bias. Consider, for example, the measurement of disability after stroke. Disabled survivors are more likely to move to a nursing home than able bodied survivors and so may also be more difficult to find and survey. A measurement of the prevalence of disability based only upon the responders might therefore be an underestimate. Worse, if the causes of non-response are associated with the outcome *and* the hospital which treated the patient, the comparison of outcome between hospitals – as, for example, in the measurement of quality of care - might also be biased (Hennekens and Buring 1987). Thus, for example, a comparison of disability after stroke might be biased if the hospitals differ substantially in the proportion of disabled patients discharged to a nursing home. Unfortunately, it is difficult to know exactly which factors are associated with non-response and even more difficult to know whether they are associated with the outcomes and hospitals under investigation. From a practical perspective, therefore, the only sure way to prevent non-response bias is to keep non-response to a minimum. The second reason to avoid non-response is statistical. As

noted (see section 1.5.3), a major problem in using outcomes to measure the quality of care is that, at most hospitals, the numbers of patients admitted with stroke over one year is relatively small. Comparisons of outcome between hospitals are therefore often imprecise. If non-response further reduces the effective size of hospital samples, the probability of identifying real differences in outcomes is also reduced. High response rates therefore also improve the confidence in comparisons of outcome.

After the first ten months of collecting outcome data in the SOP, it was clear that about a third of patients discharged from hospital had not responded to our outcome questionnaire. At this level of non-response, the possibility of bias in our estimates of functional outcome appeared strong and reduced power in their comparison a fact. The purpose of the studies described in this chapter were two-fold. First, I wished to investigate methods to improve the response to our outcome questionnaire, keeping in mind that the methods should be applicable to a routine system of follow up i.e. that they should be practical and simple; and second, I wished to determine how non-response affected our estimates and comparisons of outcome.

6.2 Improving response to the outcome questionnaire

6.2.1 Study rationale

I tested two methods designed to improve response.

1. Mailing the outcome questionnaire directly to the patient

Our original system of follow up (described in section 3.8.2) used the GP as a middle-man to circumvent the lack of detailed patient address data on the SMR1 and to prevent follow up in recently deceased cases. However, using this approach, a potentially important reason for non-response might be failure of GPs to send on our outcome questionnaire either because inadequacies in our SMR1 GP address data or because of simple lack of co-operation. Directly mailing follow up to the patient after obtaining their survival status and full address data from an alternative source might avoid these difficulties. A direct system like this is in fact a realistic possibility: the updated version of the SMR1 has the capability of reporting the patient's entire address rather than simply their post code while the General Register Office (GRO) maintains a computerised record of all deaths with new cases entered within three weeks. The aim of this sub-study was to test this more direct system.

2. Sending a second outcome questionnaire to non-responders

It is well known that response to mailed follow up can be improved by sending further follow ups to non-responders (Scott 1961; Yammarino *et al.* 1991a). Surveys that use three or four waves of follow up have been shown to achieve response rates of over 90% (Dillman 1978b). However, such intensive methods are unlikely to be

applicable in a routine system. It may also be counter productive to antagonise GPs by sending them a stream of reminders when the reason for non-response may lie with the patient. Similarly, some patients may not respond because they do not wish to dwell on their problems. Repeated requests to these patients might be viewed as unethical. In a routine setting, therefore, it seems unlikely that more than one reminder would be feasible or acceptable. Hence, the aim of this sub-study was to test the impact of a single reminder questionnaire.

6.2.2 Methods

Study design

I compared the response to the original, 'indirect' method of follow up with that to the new, 'direct' method of follow up by means of a randomised, controlled trial (RCT). I investigated the impact of sending a reminder by comparing the response rate before and after mailing the second follow up. All patients listed for follow up in the SOP were eligible for the study. Because of the nature of SMR1 data, at the time of randomisation we could not definitively identify completed hospital stays, a patient with stroke or a patient alive at the time of follow up. This caused some difficulty with randomisation and the selection of the outcome measures (see below).

Follow up procedure

The direct and indirect systems used very similar covering letters for both the GP and patient (see Appendix 10) and our standard (identical) outcome questionnaires

(Appendix 4). Patients were randomised six weeks before follow up was due and follow up sent three weeks later.

Direct follow up

I obtained the following data for each patient between randomisation and mailing:

- survival status by weekly link to GRO death certificate data
- pre-admission address from the hospital medical records office (to mimic a more detailed SMR1)
- passive consent from the patient's GP by sending a letter explaining our intent to contact the patient and asking the GP to inform us *only* if we should *not* do so (the letter included a pre-printed form and a freepost envelope for this purpose; see Appendix 10).

We mailed the follow up pack to the patient if, according to GRO, they were still alive and if the GP did not withhold their consent.

Indirect follow up

I used the standard follow up described in section 3.8.2 except for one modification. While checking the survival status of the patients in the direct group with GRO, I took the opportunity to do the same with the patients in the indirect group. Since survival at six months is in any case measured by linkage to death certification data (see section 3.8.1), the check with GRO served simply to reduce the number of

follow ups mailed. The 'middle-man' function of the GP for surviving patients was unaltered.

Non-response and sending a second follow up

I defined initial non-response as the failure to *return* a questionnaire to the trial office in time to stop a second follow up being sent at 28 days after the first mailing. For the direct method, I sent the reminder without rechecking with GRO and the GP. The covering letters to the GP and patient emphasised the importance of the study and of identifying the outcome of non-responders. I defined final non-response as the failure to *return* an outcome questionnaire by 63 days after initial mailing (35 days after sending the reminder).

Incorrect GP or patient address

Prior to the trial, I noted that GPs sometimes replied that the patient was not known to their surgery or had now changed to a different surgery. I reasoned that a routine system might have the facility to link with the database of the relevant Health Board, identify the correct or new GP, and re-direct the follow up. Similarly, where a letter sent directly to the patient's home was returned by the Post Office as 'not known at this address', a link with the Health Board might allow the new GP and patient's address to be identified, and the follow up to be re-directed. I investigated the benefit of these strategies in both arms of the trial.

Outcome measures

The primary outcome measures were

- the proportion of cases either dead at follow up or who responded within 63 days.
- the difference in the proportion of cases 'dead at follow up or who responded' at 28 days (before sending a reminder) and at 63 days (after sending a reminder).

The primary outcome measures allow an 'intention to treat' analysis (a comparison of all patients entered into the trial) and so include some patients who, after auditing the medical record, were found not to have had an acute stroke. To account for the fact that some patients had died by the time of follow up, the primary outcomes also had to include death. However, my key interest was in patients with acute stroke and, in particular, patients with acute stroke alive at the time of follow up. I therefore examined the response from these groups as secondary outcomes. Unfortunately, these comparisons are not between truly randomly allocated groups.

Randomisation

I used a computer generated minimisation technique to allocate patients to follow up. I used sex and age (under 65, 65 to 74, 75 years and above) as stratification factors. Ideally, I would like to have randomised patients once only per hospital stay. However, as described in section 3.4.2, because our database sometimes listed the same hospital stay using more than one SMR1, a small proportion of patients were allocated to receive more than one follow up. In some of these cases, both methods of follow up were used; for analysis, I have assigned them to the *first* method used.

Sample size

I estimated that the indirect method, without the addition of a reminder, would result in a similar level of response from patients discharged alive as the method employed before the trial i.e. about 65%. I estimated that the addition of a reminder would increase response by about 10% i.e. from 65 to 75%. I assumed that it would be worthwhile looking for an improvement above this level of response of about 10% e.g. 75% response in the indirect method and about 85% (or more) in the direct method. I designed the trial to have 90% power to reliably detect at least an 10% absolute difference in response above a baseline of 75% when the null hypothesis is rejected at p values of 0.05 and below. This required a total of 708 patients divided equally between the two groups. I calculated the sample size requirements using Epi Info (version 6.04b).

Ethical approval

I obtained ethical approval for the trial from the Lothian Research Ethics Committee see Appendix 11).

6.2.4 Results

Randomisation

846 cases on the study database were entered into the trial. Of these, 829 referred to a single hospital stay (416 direct and 413 indirect). Seven cases in the direct arm and eight in the indirect arm of the trial were listed more than once for the same hospital

stay (Table 6.1). Four were allocated to *both* methods of follow up on the same day; to retain these cases, I randomly assigned two to each arm of the trial.

Baseline characteristics

Of the 829 cases, 71% (588) had an acute stroke, 3% (25) had an ineligible or SAH stroke, 20% (166) had not had a stroke; and 6% (52) had no diagnosis because of failure to find the medical records or because of lack of diagnostic data therein. There were only minimal differences in the age, sex, socio-economic status, hospital of admission and proportion with an acute stroke between the two groups (Table 6.2). The subgroups of patients with acute stroke were also similar in terms of stroke pathological subtype, prevalence of ischaemic heart disease and diabetes mellitus, the proportion discharged to their pre-admission address and the proportion still in hospital five months after admission (Table 6.3). However, in terms of our key predictive variables (lived alone pre-stroke, functional independence pre-stroke and stroke severity) patients with acute stroke in the indirect group had a worse profile than those in the direct group (Table 6.3). As a result, more patients died by six months in the indirect arm (17% overall, 15% of acute strokes) than in the direct arm (12% overall, 10% of acute strokes).

Comparison of the indirect and direct methods of follow up

The results of the trial are shown in Table 6.4. There was no significant difference between the two methods of follow up in terms of the primary outcome nor in terms of the secondary outcomes i.e. both the indirect and direct methods of follow up

appeared equally effective at eliciting a response from patients who, by routine discharge data, *might* have had an acute stroke and from patients who, after audit of the medical record, were *known* to have an acute stroke and who were alive at the time of follow up.

The impact of sending a second follow up to non-responders at 28 days

The addition of a second follow up considerably increased the response in both arms of the trial (Table 6.4). The proportion with a measured primary outcome (the proportion dead or who responded by day 63) improved from 65-67% to 83-84%, an *absolute* increase of 16-19% and a *relative* increase of 24-29%. Similar sized improvements occurred in the secondary outcomes.

Response after a single follow up was non-significantly lower using the indirect system. This difference reflected the definition of response as the *return* of the questionnaire and the inherent delay in contacting patients in the indirect system. When response was defined as the date the patient *completed* the questionnaire, it was clear that (for patients with an acute stroke alive at six months) response by day 28 was the same using either system (67% direct; 64% indirect; see Figure 6.1).

The benefit of linking with the Health Board

Overall, we obtained new address details from the Health Board in 24 (3%) of the 829 patients, 12 in each arm. As a result of mailing to the new address, we obtained six questionnaire returns in the direct arm and eight in the indirect arm. Linking with

the Health Board therefore increased overall response in terms of the primary outcome by 2% (14 of the 691 cases dead or responders by day 63).

6.3 Investigating the potential for non-response bias

6.3.1 Background

To determine whether non-response is a source of bias, it is first necessary to determine whether the *outcomes* of the non-responders and responders are systematically different. Indirect means are often used for this purpose: either the baseline characteristics of responders and non-responders are compared or the outcome of early and late responders are compared (Scott 1961; Sheikh and Mattingly 1981). Ideally, however, the outcome of the non-responders should be surveyed directly. I used this method to determine the outcome of the final non-responders within the follow up trial. To determine whether any differences between responders and non-responders were sufficient to substantially bias the estimates and comparisons of outcome, I compared the outcomes of responders (to one or to two postal follow ups) with the outcomes of the entire cohort. I limited this analysis to the sample of 512 patients with an acute stroke alive at the time of follow up. The study was approved by the ethics committee at Hospital D.

6.3.2 Methods

I defined return of the outcome questionnaire by day 28 'early response'; by day 63 'late response' and failure to return any questionnaire as 'non-response'. I attempted to contact non-responders first by contacting their GP to ascertain survival status, permission to make contact and current address. If the patient was no longer registered with the GP, I identified the new GP from the Health Board; if the Health Board did not know the patient's current GP or address, I attempted to identify the data from the hospital's information system.

I attempted to contact non-responders by telephone. When a patient could not come to the phone, or if proper communication was not possible, I interviewed a proxy. I sent a further postal questionnaire if the patient could not be contacted by phone or if it was requested. I interviewed the patient in person only if it was necessary and feasible.

I took a standard approach to each telephone interview. I prefaced each with an explanation of the study and the importance of identifying the outcome of non-responders. I reassured the patient or proxy that they were not obligated to participate and that they did not have to answer every question. I read the questions out exactly as they were written on the outcome questionnaire and always in the same order. As far as possible, I used standardised prompts and explanations.

For the purpose of this sub-study, I have reported the following outcomes: functional status (by simple questions and MRS), current place of residence (by the SOP residence question) and response to the Euroquol quality of life questions (the walking, self care, usual activities, pain/discomfort and anxiety/depression domains). I have also reported the reasons for non-response.

6.3.3 Results

Identifying non-responders

Of the 512 stroke patients alive at six months, 320 were early responders, 104 were late responders and 88 were non-responders (Figure 6.2). I was able to obtain outcome data in 75% (66) of the non-responders.

Of the 22 un-traced non-responders, contact with the GP revealed that six had died *after* six month follow up was due. In a further case, the GP revealed that the patient had died shortly after hospital admission i.e. both our SMR1 and our 'gold standard' survival data - record linkage performed in 1999 – were wrong. The remaining 15 un-traced non-responders were alive. In six, the GP said it was inappropriate to make contact (two of these patients died shortly after); in two cases, I could not locate the correct GP; one patient did not wish to participate; one failed to reply despite making contact; and five were not followed in error. The predicted risk of 'death or dependency' (by our study predictive models) of the traced non-responders was not significantly different to that of the un-traced non-responders (70% vs. 49%, p 0.519).

The methods used to contact the 66 non-responders are shown in Table 6.5. Overall, I obtained outcome data by telephone in 86% and collected data directly from patients in 41% and from proxies in 59%. In five cases the only proxy I could find was a GP. In two patients response was limited to identifying their current residence.

Comparison of outcomes between early, late and non-responders

The outcomes of the early, late and non-responders are compared in Table 6.6. A greater proportion of late and non-responders were dependent when measured using the simple questions than early responders (70-72% vs. 66%) but this difference was not significant. However, there were significant trends for later responders to live somewhere other than in their own home (in a residential or nursing home, hospital or some other place); to not live with their family; to be less likely to be able to walk and to perform their usual activities. There was also a trend for later responders to be less likely to be able to wash and dress themselves which just failed to reach significance (p 0.08). There was no significant association between time of response and reports of pain & discomfort or of anxiety & depression.

It was surprising that later response was significantly associated with greater difficulty in walking, usual activities, washing and dressing (probably) and the need to live away from home and yet not with greater dependency. I hypothesised that this might relate to the measurement properties of the simple dependency question. To investigate, I compared the dependency data obtained using the simple question with that derived using the MRS (a 'gold standard'; analysis restricted to 389 patients

with both types of dependency data) and also tabulated dependency data defined by the MRS against time of response (analysis includes all 399 patients with MRS data).

The simple dependency question and the MRS disagreed on dependency in 11% of cases (Table 6.8a). The principal reason for disagreement was that, in about a fifth of cases, the simple dependency question was 'over sensitive' and tended to label a proportion of functionally impaired but independent patients as dependent (the simple question was 96% sensitive but only 78% specific). Using the MRS to define dependence, the expected trend for a greater prevalence of dependency in late and non-responders was found (p 0.012; Table 6.8b). The proportion of blank or mis-filled responses was not significantly different between the two methods (MRS: 18 cases (5.1%); simple dependency question: 12 cases (3.4%) p 0.54 (analysis restricted to the 352 patients with both sets of dependency data *and* who responded prior to the telephone survey).

Comparison of outcomes between early responders and the whole cohort

The outcomes of the early responders (i.e. estimates of outcome based on about a 65% response rate) and of the whole cohort are compared in Table 6.7 and in Table 6.8b. These analyses show that the estimates of outcome based upon the findings of the early responders were generally over-optimistic. In absolute terms, the errors were modest (4% or less) and none were statistically significant. The only exception was dependency as measured by the MRS where the finding in the early responders

was, in absolute terms, a 6% under-estimate (57% vs. 63% dependent), but this too was not significant.

Comparison of outcome between hospitals stratified by response status

Tables 6.9a and 6.9b show comparisons *between* hospitals of the proportion of patients dependent in ADL and the proportion living at home, stratified by the number of follow ups sent. For these analyses, I defined dependency using the MRS (dependency defined as a MRS score of 3-5) and residence 'at home' using the SOP method (institutional care vs. other; see section 3.7.3)

Taking the outcomes of the entire cohort as the gold standard, the comparison based upon early response alone overestimated the differences in dependency between hospitals by 3 to 22%; this error was smaller when early and late responses were combined (2 to 11%). For the proportion living at home, the comparisons based upon early response and on early and late response combined were only minimally in error (0 to 5% and 2 to 7%), respectively. In all analyses, the point estimates of the relative risk ratios based upon early response alone fell within the 95% confidence interval of the true relative risk ratios i.e. the comparisons based upon early response alone were not significantly biased.

Reasons for non-response

I identified a reason for non-response in 47 of the 66 final non-responders (in the remaining 19 cases the patient could not give a clear reason or, because postal follow

up was used, no reason was requested) (see Table 6.10). Two reasons accounted for 83% of non-response: either the respondent could not remember ever having received the questionnaire or they had received it but, for various reasons, had failed to fill or return it. Nearly twice as many in the direct arm could not remember receiving the questionnaire as in the indirect arm (36 vs. 20%).

6.4 Discussion

This study confirms our previous observation that only about two-thirds of patients with a cerebrovascular disease code listed as the primary SMR1 diagnosis - the type of patients likely to be included in a routine system for measuring the quality of stroke care - are likely to respond to a single postal follow up. The same finding applies to the sub-group of patients with acute stoke alive at the time of follow up. At this level of response, it is unlikely that a routine system of measurement would be credible because of reduced power and the potential for bias in the comparisons of outcome.

This study helps to quantify the potential for bias. Although non-responders to a single follow up were significantly more likely than responders to be restricted in ADL and to require institutional care, these differences were small in absolute terms and hence, *as a group*, the outcomes of the initial responders were not significantly different to the outcomes of the entire cohort. Even after a single follow up, therefore, error in estimates of dependency in ADL and in the need for institutional care seem likely to be modest at the majority of hospitals. However, the potential for

biased comparison of outcome, in this case dependency in ADL, appears to remain between *some* hospitals. Thus, after a single follow up, the estimate of the difference in dependency between Hospital A and Hospital B was substantially wrong (22% overestimate; true relative risk ratio 1.19; estimated relative risk ratio 1.41), a level of error that, if real, would certainly mislead a study which aimed to use dependency data to indicate differences in the quality of care. The use of a second follow up led to a substantial increase in response (up to 83-84%). At this level of response, the potential for error in the comparisons of dependency halved but continued to vary between hospitals (from 2 to 11%). Although undoubtedly an improvement, the potential for moderately biased comparison of dependency between some hospitals would appear to remain – after all, an 11% difference in dependency approximates to about a third of the treatment effect of one hospital admitting *all* its patients to a stroke unit and the other hospital admitting *none* (Stroke Unit Trialists' Collaboration 1997a). That said, the precise interpretation of all these errors is difficult because the small size of the hospital samples means that none were significant i.e. they may quite plausibly be due to chance rather than bias. Clearly, however, the potential for important bias cannot be excluded.

My findings relating to response rate and bias are likely to be generalisable given that they derive from a consecutive series of patients discharged from five, typical Scottish hospitals; that the levels of response are in good agreement with other studies that have reported response to two waves of postal follow up after stroke (Dorman *et al.* 1997a; Parker *et al.* 2000); that others have found response to postal follow up after stroke to vary substantially between hospitals (Parker *et al.* 2000);

and that some (Hoeymans *et al.* 1998), although not all (Parker *et al.* 2000), have also shown that the prevalence of disability is higher in non-responders than in responders to a single postal contact. Taken together, they strongly suggest that a *routine* system aiming to follow stroke patients several months after discharge should, as a minimum, have the facility to send a second follow up to initial non-responders. However, as noted, even with such a system, misleading comparisons of dependency might still result. Fortunately, as will be shown, other routinely practical strategies might also be used to improve response and reduce non-response bias to an acceptable level.

Contrary to expectations, response was not improved by excluding the GP from the system of follow up. There may even have been a small disadvantage in doing so since response by the indirect method was marginally greater than by the direct method and because this difference may have been a slight under-estimate (the indirect arm of the trial included more patients with baseline characteristics indicative of severe stroke i.e. those less likely to respond). Any benefit from retaining the GP most likely relates to their knowledge of their patient's current whereabouts since nearly twice as many final non-responders in the direct arm could not remember receiving our questionnaire. In the absence of a major difference in response, the choice of method of follow up for a routine system should be guided by considerations of cost and acceptability. The indirect method is simple, cheap and immediately feasible. However, whether GPs would find it acceptable to act as a middle-man to a real system is unknown, especially as such a system would need to accommodate reminders and perhaps also patients with a number of disorders other

than stroke. Moreover, there is evidence that GPs are already reluctant to participate in postal questionnaire surveys (Barclay *et al.* 2002). Although more complicated and expensive to establish, the direct system might also prove cheap to run and, by minimising work for GPs and their staff, might be more acceptable in the longer term. Regardless of the method chosen, it is clear that contact with the Health Board only marginally improves overall response and hence would be worthwhile only if a routine link were simple and inexpensive.

Follow up of survivors of stroke is challenging given that, as group, they are likely to experience the problems of communication associated with older age (cognitive impairment, poor vision, deafness) and those associated with stroke (dysphasia, dysarthria, disability and depression). Furthermore, a substantial proportion move or to remain in hospital by six months after admission and hence are more difficult to find. In this context, a five-sixths response to an unheralded postal follow up (plus one reminder) is actually very good. Nonetheless, our experience suggests that other simple strategies might improve response further.

First, it might be possible to improve the delivery of the outcome questionnaire. Both systems of follow up had to deal with the fact that the name of the patient's GP is an optional field on the SMR1 and hence is often not reported. In these cases, we had to mail follow up to the address of the surgery alone. This was problematic when several surgeries shared the same building (as occurred in some urban areas, especially at Hospital B) or where a surgery had two or more sites. Staff had

difficulty in identifying the relevant GP and we received a number of letters complaining about the work generated in doing so. Mandatory reporting of the GP's name on the SMR1 would have prevented this problem. The systems of follow up also had to deal with the fact that the SMR1 reports the patient's address *prior* to admission. It is common-sense to suggest that improved response from patients who change their place of residence after a stroke might follow if the SMR1 were to report the address to which the patient was *discharged.* In this regard, it is worth recalling that the Health Board maintains a central database listing the patient's current GP *and* address and it is conceivable that, if ever set up, an official system of follow up might be able to automatically link with these data. Co-operation from GPs and response from patients might also be higher if a national system of follow up were known to be in operation (Sloan *et al.* 1997). For an indirect system, a covering letter making clear that the follow up had come via their GP might also help (Smith *et al.* 1985).

Second, it might be possible to improve response by improving the format of our questionnaire. Thus, our questionnaire was rather plain, printed on white paper and, including the MRS, was four (single-sided) pages long (see Appendix 4). A more professional, shorter and possibly larger print version (perhaps asking the key questions of dependency and residence only) might be more appealing and so increase response (Yammarino *et al.* 1991b). Further attempts at follow up, perhaps by telephone, would also be likely to improve response. Telephone follow up appears to be less problematic in the elderly than was formerly the case (Wilson and Roe 1998) and the simple questions and MRS are valid when administered in this

way (Candelise *et al.* 1990a; Lindley *et al.* 1994b). However, it should be stressed that telephone follow up can involve a number of calls simply to establish contact and then sometimes a further call or postal follow up to obtain the outcome data i.e. it is labour intensive. It may not, therefore, be cost-effective.

Lastly, and importantly, the impact of variation in response to follow up between hospitals may be less serious when comparisons of outcome are adjusted for important differences in casemix. Thus, outcomes that have been adjusted for important casemix have, by definition, already been internally standardised i.e. judged to be better, worse or in-line with the predicted outcome derived from a prognostic model. As a result, comparisons of the *adjusted* outcomes are, to some extent, unaffected by the exclusion of a proportion of cases from the analysis. To illustrate, consider if only half the patients dependent at six months reply to follow up. Provided the quality of care afforded to the responding and non-responding dependent patients was the same (see section 4.4), the ratio of observed and predicted outcome in each group will be the same, which is to say, the adjusted outcome derived from the responding patients will not be biased. The problem with this assumption, however, is that worse adjusted outcome may be associated both with worse care and with non-response (e.g. the less severely dependent patients may be more likely to respond while the more severely dependent patients may not and they may have had worse care). Thus, the extent to which adjustment 'protects' against differential non-response depends on the extent to which non-responders received the same quality of care as the responders, and clearly this may vary between hospitals.

The failure of the simple dependency question to recognise a proportion of impaired but functionally independent patients, as defined by the MRS, is perhaps not a surprise. This finding reflects the trade off between simplicity (and hence ease of completion) and detail (and hence fine discrimination) accepted in its development (Lindley *et al.* 1994b). The simple dependency question was originally developed for use in treatment trials with many thousands of patients in each arm where less refined measurement of outcome is not essential (Lindley *et al.* 1994b). However, its lack of refinement may be a problem in studies with only a few hundred patients in each group (i.e. in studies of quality of care) where it appears to obscure real differences in functional outcome (and so, *a priori*, reduces the ability to differentiate between hospitals with different standards of care). Moreover, the proportion of blank or mis-filled responses was small and not significantly different with the simple dependency question and with the MRS. Given its greater refinement, it is arguable, that the MRS should be the preferred measure of dependency for the purpose of routinely measuring the quality of stroke care.

Indeed, given the small sample sizes and potential for dilution of treatment effects (by the inclusion of patients too well or unwell to show appreciable benefit), the case for a more discriminating measure of functional outcome in studies of the quality of stroke care can be extended. The dichotomisation of the MRS into groups with 'good' or 'bad' functional outcome (e.g. MRS 0-2 vs. 3-5) inevitably leads to a loss of information and hence to less powerful comparisons. A more discriminating strategy might be to abandon the concept of dependency and instead to compare responses across the entire six point scale of the MRS using non-parametric

statistical methods (Duncan *et al.* 2000). Re-analysis of the Second European-Australasian Acute Stroke Study (ECASS 2) of thrombolysis after ischaemic stroke illustrates this point (Stingele *et al.* 2001). Using the prespecified primary endpoint of the percentage of patients with favourable outcome (MRS scores zero and one) the trial was inconclusive, the 3.7% absolute improvement with thrombolysis over placebo failing to reach statistical significance (p 0.277). However, using non-parametric methods to compare the entire MRS outcome distributions between the two groups, it is clear that the outcomes of patients treated with thrombolysis were just significantly better (p 0.047).

A better approach still might be to use more finely scaled but still relatively simple instruments that measure ability in terms of basic and extended ADL (Duncan *et al.* 2000), for example, the combination of the Barthel Index and the Nottingham extended ADL index (Wade 1995b), or the recently (and rigorously) developed Stroke Impact Scale 16 (SIS-16) which incorporates both aspects of ADL, hand function and mobility in a 16 point measurement tool (Duncan *et al.* 2003). Importantly, the Stroke Impact Scale (from which the SIS-16 is derived) has been shown to be valid when administered by postal survey (Duncan *et al.* 2002b) and by proxy (Duncan *et al.* 2002c). Of course, the argument for more discriminating measures can be extended even further to include measures of cognition, communication, and higher social function. However, longer measures may be counter-productive in terms of greater non- and incomplete response, especially in an elderly and brain damaged population and in a routine setting where the ability to make repeated contact and to query incomplete or mis-filled responses is limited

(Parker *et al.* 2000). For routine systems which aim to measure the quality of stroke care a compromise between the detail and the practicality of the outcome measure is clearly needed. While the suggested alternatives to the MRS may be theoretically attractive, their practical suitability should be tested before any superiority in a *routine postal* setting is automatically assumed.

The studies described in this chapter have a number of methodological shortcomings. Fortunately, these are unlikely to be a major source of bias or were largely unavoidable. First, the unit of randomisation in the trial was not the unit of analysis in all cases. However, this problem affected only a small and equal proportion (2%) of cases in each arm of the trial. Second, the trial did not primarily study patients with acute stroke and, because of limited information on the SMR1, could not stratify by the clinical baseline characteristics which influence response. The practical result is that I may have modestly under-estimated the response to the indirect system of follow up. A benefit, however, is that the findings are generalisable to the population of patients likely to be included in any real system of follow up. Third, I failed to contact 22 (25%) patients in the survey of non-responders. However, this was unavoidable in seven cases (because the patient had died) and the traced and un-traced non-responders had similar predicted risks of being dead or dependent at six months. Fourth, the method used to survey responders (postal questionnaire in 100%) differed from that used to survey non-responders (telephone in 86%). In mitigation, the simple dependency question has been shown to be equally accurate when administered by telephone or by post (Lindley *et al.* 1994a); and Rankin scores obtained by face-to-face interview agree with those obtained by telephone (Candelise

et al. 1990b). Fifth, I accepted the GP as a proxy in the survey of non-responders. A GP's estimate of functional status might not be as valid as that given by a partner or regular carer. However, this potential bias occurred only in five of the 66 cases surveyed and I used the GP as a proxy only if they had seen the patient recently. Lastly, the study involved too few hospitals and too few patients to definitively determine whether non-response leads to bias in the comparison of dependency between some hospitals. Ideally, a much larger study involving more hospitals and several hundreds of patients at each should be performed to confirm or refute this possibility.

Summary

1. Response to a single, unheralded postal follow up six months after admission for acute stroke was in the region of 60-65%. The addition of a second follow up increased response to 83-84%.

2. Later response to follow up was associated with a significant trend toward greater restriction or dependency in ADL and greater likelihood of institutional care. However, in absolute terms, the differences in outcome between early and later responders were small.

3. Even after a single follow up, estimates and comparisons of dependency and institutionalisation were reasonably accurate at many hospitals. However, at some hospitals, comparisons of these outcomes may be unacceptably biased even after a second follow up.

4. As a minimum, a routine system of follow up should have the facility to send a second follow up to initial non-responders. However, to reduce non-response bias to an acceptable level, other strategies should be employed.

5. Response is not increased by posting follow up directly to the patient's pre-admission address. The choice between a direct or indirect system of follow up should be made on grounds of cost and acceptability.

6. Improved delivery of the outcome questionnaire, and hence improved response, might result if the SMR1 routinely listed the name of the patient's GP and the address to which the patient was discharged. A simpler and more authoritative questionnaire might also improve response. Further attempts at follow up, in particular by telephone, would also improve response but might not be cost-effective.

7. Adjustment for casemix reduces bias due to non-response provided the responders and non-responders have received care of the same quality. However, by the very nature of non-response, this may not always be the case and hence the degree of 'protection' afforded by adjustment may vary between hospitals.

8. The MRS may be better suited to the measurement of dependency than the simple dependency question for the purpose of measuring the quality of stroke care. Comparison of functional outcome across the entire range of the MRS rather than its simple dichotomisation is likely to be more sensitive to differences in quality of care. The suitability of more detailed measures of functional outcome to a routine setting merits investigation.

Table 6.1 Number of times a single hospital stay was allocated to follow up

Number of times a single stay was allocated	Direct follow up	Indirect follow up
Once	409	404
Twice	7	8
Thrice	0	1

Table 6.2 Baseline characteristics of the patients in the follow up trial

	Direct follow up		Indirect follow up		p (χ^2)
	n	%	n	%	
Patients	416		413		
Median age (in years)		**74**		**73**	0.97*
Male	196	**48**	197	**48**	0.91
DepCat† 1 to 2	46	**11**	49	**12**	0.74
3 to 5	262	**64**	250	**61**	
6 to 7	103	**25**	110	**27**	
Hospital A	84	**20**	101	**25**	0.53
B	94	**23**	97	**23**	
C	70	**17**	68	**17**	
D	106	**26**	90	**22**	
E	62	**15**	57	**14**	
Audited diagnosis ‡					
Acute stroke	287	**69**	298	**72**	0.85
Other diagnoses	129	**31**	115	**28**	0.75
Ineligible or SAH stroke	11	3	14	3	
Not stroke	88	21	78	19	
Not known	30	8	23	5	

* Mann Whitney U test

† Missing values for 9 cases (5 Direct, 4 Indirect)

‡ Identified by inspection of the medical record

Table 6.3 Baseline characteristics of acute stroke patients in the follow up trial

	Direct follow up		Indirect follow up		p (χ^2)
	%	n	%	n	
Patients	**49**	287	**51**	298	
Demographics					
Median age (years)	**73**		**72**		0.57 *
Male	**49**	140	**48**	143	0.91
Deprivation Categories					
1 to 2	**11**	32†	**13**	37‡	0.49
3 to 5	**64**	183†	**60**	176‡	
6 to 7	**24**	69†	**28**	82‡	
Hospital					
A	**16**	45	**22**	64	0.27
B	**22**	62	**23**	69	
C	**19**	53	**19**	57	
D	**29**	83	**23**	69	
E	**15**	44	**13**	39	
Stroke pathological subtype					
Ischaemic stroke	**79**	227	**82**	244	0.68
Haemorrhagic stroke	**11**	32	**10**	30	
Type unknown	**10**	28	**8**	24	
Pre-stroke					
Lived alone	**36**	101†	**43**	127	0.10
Independent in ADL	**88**	251†	**90**	267	0.74
Diabetes mellitus	**13**	37§	**15**	43#	0.72
Ischaemic Heart Disease	**33**	92§	**31**	92#	0.74
On admission					
Orientated & able to speak	**72**	204†	**68**	202	0.33
Able to lift both arms	**74**	210†	**70**	209	0.35
Able to walk without help	**41**	117†	**36**	108	0.25
Incontinent of urine	**42**	117§	**42**	125‡	0.90
At discharge					
In hospital > 5 months	**2**	5	**2**	6	0.95
Returned to pre-admission address	**78**	220†	**75**	222	0.46

* Mann Whitney U test

Denominator d varies in some cases due to missing data
(† d = 284, ‡ d = 295, § d = 282, # d = 297)

Table 6.4 Response to the Direct and Indirect methods of follow up before and after sending a reminder to initial non-responders

Outcomes identified	**Follow up method**				**Relative risk ratio** (95% CI)
	Direct		**Indirect**		
	%	n	%	n	
Primary outcomes (n 829)		416		413	
Dead or responded within 28days	**67**	280	**65**	270	1.04 (0.90, 1.21)
Dead or responded within 63 days	**83**	344	**84**	347	0.95 (0.80, 1.14)
Secondary outcomes					
Acute strokes (n 585)		287		298	
Dead or responded within 28 days	**68**	195	**66**	198	1.04 (0.87, 1.23)
Dead or responded within 63 days	**83**	239	**87**	258	0.88 (0.71, 1.09)
Acute strokes alive at 6 months (n 512)		260		252	
Responded within 28 days	**65**	168	**60**	152	1.10 (0.91, 1.31)
Responded within 63 days	**82**	212	**84**	212	0.92 (0.74, 1.13)

Table 6.5 The methods used to elicit a response in the survey of patients with a stroke and alive at six months who did not return an outcome questionnaire i.e. in cases who were non-responders by day 63.

Respondent	Telephone interview	Postal questionnaire	Face to face interview	Total
Patient	23	3	1	27 (41%)
Relative	16	0	0	16 (24%)
Carer	12	4	0	16 (24%)
Friend	1	1	0	2 (3%)
GP	5	0	0	5 (8%)
Total	57 (86%)	8 (12%)	1 (2%)	66

Table 6.6 Comparison of outcomes for early, late and non-responders

Outcome		Early responders n = 320		Late responders n = 104		Non-responders n = 66		
		n	% *	n	% *	n	% *	p (χ^2 trend) *
Dependent in everyday activities	Yes	208	**66**	73	**72**	45	**70**	0.36
	No	106		29		19		
	Missing	6		2		2		
Left with any problems	Yes	254	**80**	80	**79**	51	**81**	0.35
	No	48		21		12		
	Missing	18		3		3		
Now living at home ‡	Yes	254	**80**	76	**75**	40	**61**	0.001
	No	63		25		26		
	Missing	3		3		0		
Now living with family ‡	Yes	54	**17**	12	**12**	1	**2**	0.0009
	No	263		89		65		
	Missing	2		1		0		
Now living in a residential or nursing home ‡	Yes	24	**8**	13	**13**	13	**25**	0.002
	No	293		88		53		
	Missing	3		3		0		
Now living in a hospital or other place ‡	Yes	23	**7**	10	**10**	12	**18**	0.008
	No	294		91		54		
	Missing	3		3		0		

‡ residence questions were not mutually exclusive e.g. patients might answer that they lived at their pre-admission address & with family

* missing values excluded from calculation of percentages and χ^2

Table 6.6 continued

Euroquol		Early responders n = 320		Late responders n = 104		Non-responders n = 66		
		n	% *	n	% *	n	% *	p (χ^2 trend) *
Walking	No problems	67	**22**	24	**24**	15	**23**	0.006
	Some problems	233	**75**	69	**69**	42	**66**	
	Confined to bed	10	**3**	7	**7**	7	**11**	
	Missing	10		4		2		
Self care	No problems	131	**44**	29	**30**	22	**35**	0.08
	Some problems	133	**45**	53	**55**	29	**46**	
	Cannot wash & dress	34	**11**	15	**15**	12	**19**	
	Missing	22		7		3		
Usual activities	No problems	51	**17**	18	**18**	11	**18**	0.02
	Some problems	159	**51**	40	**40**	24	**38**	
	Unable to perform	99	**32**	42	**42**	28	**44**	
	Missing	11		4		3		
Pain & discomfort	None	108	**36**	34	**34**	26	**41**	0.71
	Moderate	169	**57**	55	**55**	32	**51**	
	Extreme	22	**7**	11	**11**	5	**8**	
	Missing	21		4		3		
Anxiety & depression	None	111	**38**	30	**31**	21	**33**	0.20
	Moderate	165	**55**	62	**63**	35	**54**	
	Extreme	20	**7**	6	**6**	8	**13**	
	Missing	24		6		2		
Respondent	Patient	149	**49**	42	**42**	27	**41**	0.15
	Other	158		60		39		
	Missing	13		2		0		

* missing values excluded from calculation of percentages and χ^2

† χ^2 test for trend: calculated by dichotomising outcome into none/ moderate versus extreme

Table 6.7 Comparison of outcomes for early and all responders

Outcome		Early responders n = 320		All cases n = 490		
		n	% *	n	% *	p (χ^2) *
Dependent in everyday activities	Yes	208	**66**	326	**68**	0.62
	No	106		154		
	Missing	6		10		
Left with any problems	Yes	254	**80**	385	**83**	0.59
	No	48		81		
	Missing	18		24		
Now living at home ‡	Yes	254	**80**	370	**76**	0.22
	No	63		114		
	Missing	3		6		
Now living with family ‡	Yes	54	**17**	67	**14**	0.22
	No	263		417		
	Missing	2		3		
Now living in a residential or nursing home ‡	Yes	24	**8**	50	**10**	0.19
	No	293		434		
	Missing	3		6		
Now living in a hospital or other place ‡	Yes	23	**7**	45	**9**	0.31
	No	294		439		
	Missing	3		6		

‡ residence questions were not mutually exclusive e.g. patients might answer that they lived at their pre-admission address *and* with family

* missing values excluded from calculation of percentages and χ^2

Table 6.7 continued

Euroquol		Early responders n = 320		All cases n = 490		
		n	% *	n	% *	p (χ^2) *
Walking	No problems	67	**22**	106	**22**	0.43
	Some problems	233	**75**	344	**73**	
	Confined to bed	10	**3**	24	**5**	
	Missing	10		16		
Self care	No problems	131	**44**	182	**40**	0.47
	Some problems	133	**45**	215	**47**	
	Cannot wash & dress	34	**11**	61	**13**	
	Missing	22		32		
Usual activities	No problems	51	**17**	80	**17**	0.48
	Some problems	159	**51**	223	**47**	
	Unable to perform	99	**32**	169	**36**	
	Missing	11		18		
Pain & discomfort	None	108	**36**	168	**36**	0.90
	Moderate	169	**57**	256	**56**	
	Extreme	22	**7**	38	**8**	
	Missing	21		28		
Anxiety & depression	None	111	**38**	162	**35**	0.82
	Moderate	165	**55**	262	**57**	
	Extreme	20	**7**	34	**8**	
	Missing	24		32		
Respondent	Patient	149	**49**	218	**46**	0.47
	Other	158		257		
	Missing	13		15		

* missing values excluded from calculation of percentages and χ^2

Table 6.8a Accuracy of dependency data obtained using the simple dependency question in comparison with that obtained using the Modified Rankin Scale (MRS).

Modified Rankin Scale	**Simple dependency question**	
	Dependent	Independent
Dependent (score 0-2)	235	11
Independent (score 3-5)	32	111

Simple dependency question

Agreement with MRS	346 / 389	**89%** (95% CI: 86 – 92)
Sensitivity	235 / 246	**96%** (92 – 98)
Specificity	111 / 143	**78%** (71 – 85)

Table 6.8b Comparison of Modified Rankin Scale scores by response status

MRS score	Response status: Early		Late		Non		All cases	
	n	%*	n	%*	n	%*	n	%*
Independent								
Scores 0-2	104	**43**	27	**30**	18	**28**	149	**37**
Dependent								
Scores 3-5	141	**57**	63	**70**	46	**72**	250	**63**
Confused/blank	15		3		2		20	
MRS not measured	60		11		0		71	
Total	320		104		66		490	

* confused/blank/unmeasured values excluded from calculation of percentages and χ^2

Comparisons of proportions with MRS 0-2 vs. 3-5

between early, late and non-responders: p 0.012 (χ^2 trend)

between early responders and all cases: p 0.20 (χ^2)

Table 6.9a Comparison of proportion dependent in ADL (by the Modified Rankin Scale) by response status between hospitals

Hosp.	Dependent MRS 3-5	Early responders n = 320				Early & Late responders n = 424				All cases n = 490			
		n	% †	RR †	95% CI	N	% †	RR †	95% CI	n	% †	RR†	95% CI
Hosp A	Yes	22	54	Ref.		34	60	Ref.		42	63	Ref.	
	No	19				23				25			
	No data	17				18				19			
Hosp B	Yes	34	76	**1.41**	1.01, 1.96	51	77	**1.30**	1.01,1.66	67	74	**1.19**	0.95, 1.48
	No	11				15				23			
	No data	13				17				17			
Hosp C	Yes	32	68	**1.27**	0.90, 1.79	48	74	**1.24**	0.96,1.60	53	73	**1.16**	0.92, 1.46
	No	15				17				20			
	No data	11				15				15			
Hosp D	Yes	34	43	**0.79**	0.54, 1.16	45	44	**0.74**	0.54,1.00	54	47	**0.76**	0.58, 0.99
	No	46				57				60			
	No data	19				20				20			
Hosp E	Yes	19	59	**1.11**	0.74, 1.66	26	58	**0.97**	0.70,1.35	34	62	**0.99**	0.75, 1.30
	No	13				19				21			
	No data	15				19				20			

† Cases with no data are excluded from calculations of percentages and relative risk ratios (RR)

Table 6.9b Comparison of proportion at home (by the SOP definition) by response status between hospitals

Hosp.	At Home	Early responders n = 320				Early & Late responders n = 424				All cases n = 490			
		n	% †	RR †	95% CI	N	% †	RR †	95% CI	n	% †	RR†	95% CI
Hosp A	Yes	50	88	Ref.		65	88	Ref.		70	83	Ref.	
	No	7				9				14			
	No data	1				1				1			
Hosp B	Yes	50	91	**0.96**	0.85, 1.10	70	89	**1.01**	0.90,1.13	89	86	**0.97**	0.86, 1.10
	No	5				9				15			
	No data	4				4				4			
Hosp C	Yes	46	82	**1.07**	0.91, 1.25	66	84	**0.95**	0.84,1.08	70	81	**1.02**	0.89, 1.18
	No	10				13				16			
	No data	1				1				1			
Hosp D	Yes	85	89	**0.99**	0.88, 1.12	101	86	**0.98**	0.88,1.10	108	84	**1.00**	0.88, 1.12
	No	11				16				21			
	No data	3				5				5			
Hosp E	Yes	41	87	**1.01**	0.87, 1.16	53	84	**0.96**	0.84,1.10	61	82	**1.01**	0.88, 1.17
	No	6				10				13			
	No data	0				1				2			

† Cases with no data are excluded from calculations of percentages and relative risk ratios (RR)

Table 6.10 The reasons given by non-respondents for not having replied to our earlier attempts to obtain outcome data

Reason	Direct follow up n =36		Indirect follow up n = 30		Total n = 66	
	%	n	%	n	%	n
GP initially did not want us to contact the patient ∴ no questionnaire received	**2**	1	**7**	2	**4**	3
Thinks that they did return the questionnaire	**6**	2	**10**	3	**8**	5
Received the questionnaire but did not fill it in	**28**	10	**33**	10	**30**	20
Cannot remember ever having received the questionnaire	**36**	13	**20**	6	**29**	19
No reason obtained	**28**	10	**30**	9	**29**	19

Figure 6.1 Cumulative response rate (where response is defined as the day the outcome questionnaire is completed) in the 512 patients with an acute stroke in the follow up trial who were alive at the time of follow up

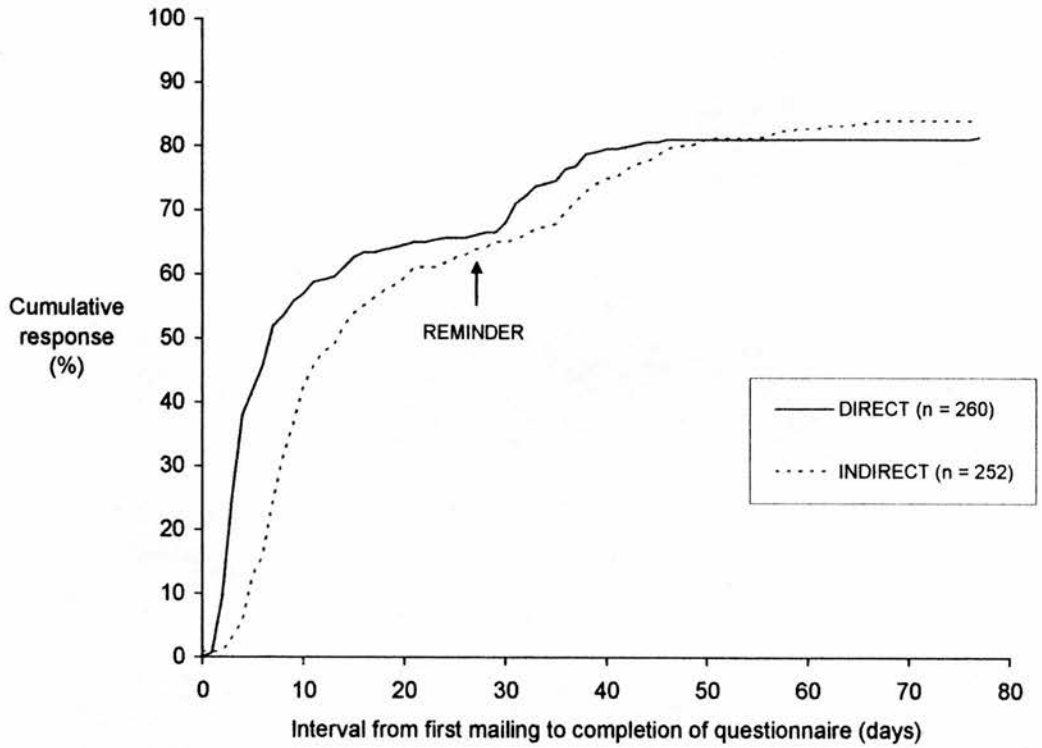
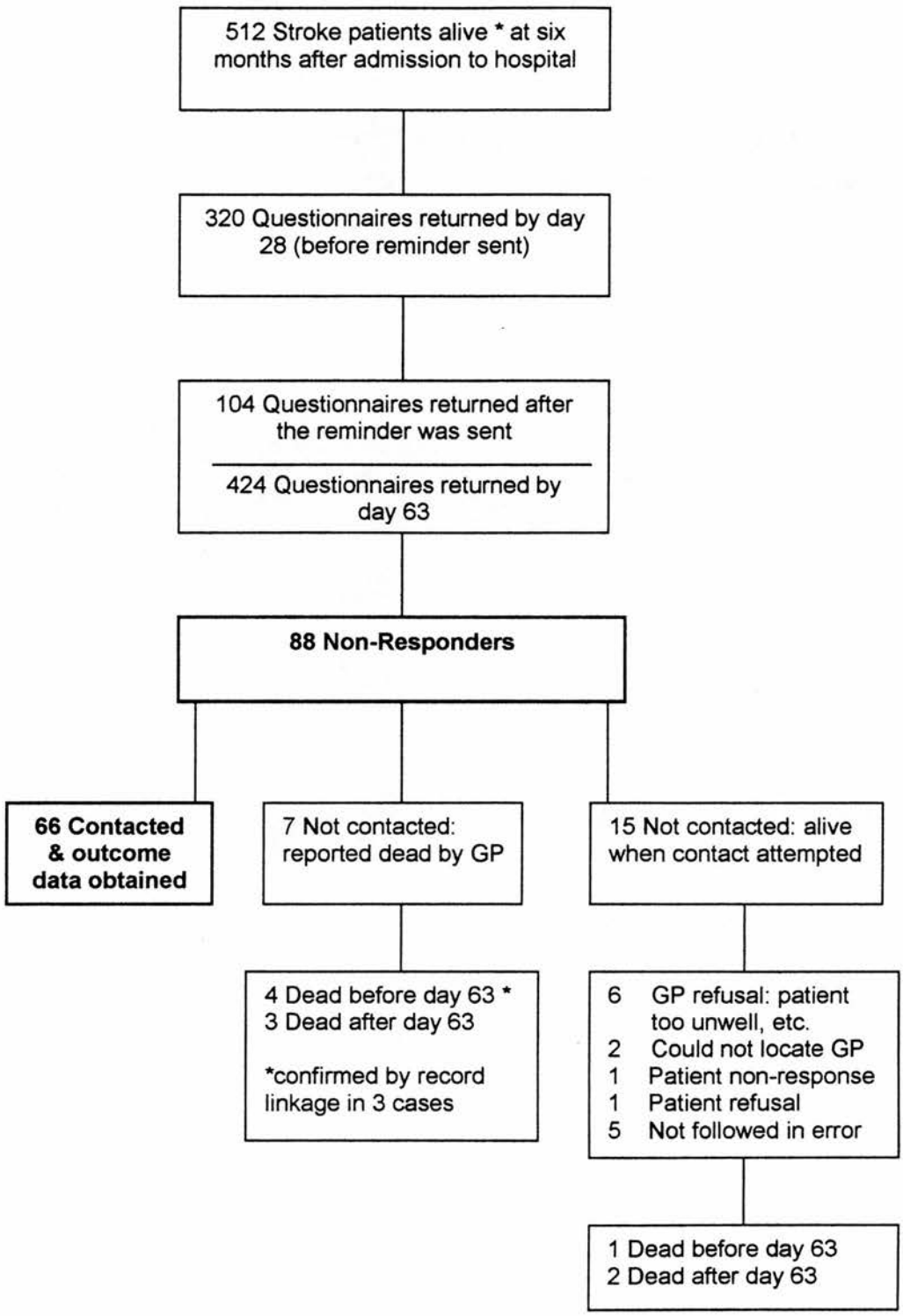


Figure 6.2 Follow up of the 512 patients with an acute stroke who were alive six months after admission to hospital



* Survival status by record linkage performed in 1999

i.e. 12 months after the end of the follow up study

Chapter Seven. SOP results (1): Outcome

7.1 Introduction

Having pondered the practicalities of routinely measuring outcome after stroke, it is now time to consider the other question addressed by the SOP, namely, whether routinely collected stroke outcomes, once adjusted for casemix (using our methods), are likely to be useful indicators of the quality of stroke care. An attempt to answer this question is laid out in this chapter and the next. This chapter describes the ascertainment of the hospital cohorts, their baseline characteristics and their outcomes before and after adjusting for casemix. Alternative methods of adjusting for casemix and explorations of the main findings are also presented. The quality of the structure and processes of stroke care at each hospital and their relationship to the adjusted outcomes are described in the chapter that follows.

7.2 Additional analyses

7.2.1 Alternative methods of adjusting outcome for casemix

In addition to our principal method of adjusting outcomes for casemix (described in section 3.9) I have also presented analyses of outcomes adjusted for:

1. Age, sex and social deprivation: to mimic the method used to adjust the stroke outcome data currently published in Scotland.

2. Urinary incontinence: to estimate whether this simple variable provides adjustment equal to that of our study models.

3. The study models *plus* urinary incontinence *or* social deprivation: to determine whether the adjustment achieved by our study models can be improved by the addition of a further simple variable. I tested the addition of urinary incontinence because it is an established predictor of outcome after stroke and social deprivation because it is routinely available in Scotland and associated with outcome after stroke (Anonymous 1997; Kunst *et al.* 1998; Modan and Wagener 1992).

4. All measured casemix variables (see Table 7.6): to determine whether any residual variation in outcome, after adjusting for the variables in our study models, might be explicable by any remaining differences in *measured* casemix.

For the first two analyses, I built logistic regression models within the SOP data set and forced in the relevant predictive variables; for the latter two, I forced in the additional predictive variables along with either the variables of the study model (for case fatality) or with the linear predictor of the study models (for 'death or dependency' and 'alive & at home'). I entered two categorical variables: social deprivation using Carstairs scores (with the least deprived social stratum as the reference category; see section 3.5.2); and the interval between stroke onset and admission to hospital (11-30 days prior (reference category), 1-10 days prior, zero days, 1-10 days after, 11-20 days after and over 21 days after admission).

7.2.2 Exploring outcomes adjusted using the study models

I performed two analyses in an attempt to understand the reasons for any residual differences in adjusted outcome between hospitals.

1. I determined how any *overall* differences in adjusted outcome between hospitals reflected the outcome of patients with mild, moderate and severe stroke (approximated by low, medium and high predicted risk of the outcome in question) by ordering the entire SOP cohort by the predicted risk, dividing it into three equal sized parts (tertiles), and then comparing the adjusted outcomes between hospitals for each tertile.
2. I determined when any differences in adjusted case fatality occurred by constructing a Cox proportional hazards regression model within the SOP data set (using the same predictive variables as in our study models) and then plotting adjusted survival curves for each hospital cohort. I checked the proportionality assumption for each variable by means of log minus log survival plots before using the model (data not shown).

7.2 Case ascertainment

The SOP identified a total of 4223 hospital admissions in which a SMR1 listed a cerebrovascular disease code as the principal diagnosis (Table 7.1). Of these, we audited the medical records of 4017 (95% overall; ranging from 93% at Hospital A to 98% at Hospital C). The main reason for not auditing admissions was inability to obtain the relevant medical records. In 10 cases the medical record did not allow us

to identify the hospital to which the patient was first admitted; I assigned these patients to hospitals using SMR1 data alone.

I compared the characteristics and outcomes of the audited and unaudited cases in Table 7.2 (using a significance level of 0.01 given the multiple comparisons). There were no significant differences in terms of age, sex, the proportion dead or the proportion dead or dependent at six months either overall or within any hospital. Un-audited cases were significantly more likely to have been coded as non-emergency admissions (p 0.01) i.e. they were less likely to have had an acute stroke than the audited cases (see section 4.3.2).

Of the 4017 audited admissions, 2845 (71%) had had an acute stroke, 160 (4%) had had an ineligible stroke (see section 3.5.1) and 1012 (25%) had not had a stroke (Table 7.3). The hospital with the highest proportion of ineligible stroke admissions (whose definition includes cases transferred from non-study hospitals) was inevitably Hospital D, the only hospital to house a regional neurology and neurosurgery centre.

The SMR1 and our audit data disagreed on the hospital to which the patient was first admitted in 17 cases of acute stroke (Table 7.4). Our audit revealed that 15 of these were actually first admitted to Hospital A or E and then later transferred to Hospital D (14 had either a SAH or haemorrhagic stroke). In these cases I assumed that our audit data were correct and altered the hospital assignment accordingly.

Of the 2845 admissions with an acute stroke, 121 (4%) were for SAH. The proportion with SAH was similar (2%) at all hospitals except Hospital D (12%). The remaining 2724 (96%) acute stroke admissions had had non-SAH strokes. The derivation of this cohort, on whom all the following analyses are based, is illustrated in Figure 7.1. The final cohort comprises 2574 patients with a single admission for acute non-SAH stroke and 73 patients with two or more admissions, each for a separate non-SAH stroke (70 had two admissions, two had three admissions and one had four admissions).

7.3 Baseline characteristics of patients with acute stroke

We extracted data describing each baseline characteristic in over 99% of admissions except for two variables (Table 7.5). These variables were whether the systolic BP was greater than 160 mmHg (data missing in 37 cases) and the interval between stroke onset and admission to hospital (exact date of onset missing in 158 cases). Our audit assistant noted particular difficulty in extracting data from the medical record in a significantly greater proportion of cases at Hospital A than at any other hospital (13% vs. 5 - 9%). A greater proportion of these cases were dead at six months at Hospital A than elsewhere (54% vs. 30 - 40%). Discussion revealed that this related, in part, to the practice of storing the medical records of deceased patients on microfiche at a much shorter interval after death at Hospital A than elsewhere. Our audit assistant commented that the medical records of these cases were often jumbled and sometimes incomplete.

We failed to collect all six study model variables in 19 cases; urinary incontinence in 14 cases; and social deprivation in 19 cases (Table 7.5). I replaced these missing values in order to retain all 2724 cases in each analysis of adjusted outcome. For the study model variables and for urinary incontinence, I replaced the missing values using the most pessimistic option e.g. for the ability to walk unaided I marked all missing values as 'unable'. For social deprivation the missing information was the post code of the patient's pre-admission address (see section 3.5.2). I first replaced any missing values using SMR1 post code data (4 cases). In the remaining cases, the post codes obtained by our audit were not listed in the Carstairs database (McCloone 1994). Here, I identified the deprivation category of any listed post codes which had the same first three symbols as the audited post code (e.g. EH1) and assigned the median value (six cases). In the remainder, I simply inserted the lowest deprivation category (nine cases). To determine whether the pessimistic replacement of missing variables was a source of bias I repeated all comparisons of adjusted outcomes with the missing values replaced using the most optimistic option.

The baseline characteristics of the patients with acute stroke (prior to the replacement of missing values) are shown in Tables 7.6a and 7.6b. Setting p to 0.01 to allow for multiple comparisons, there were highly significant differences between hospital cohorts in terms of age, social deprivation, pre-stroke employment status, interval between stroke onset and admission, proportion of strokes which occurred after admission for another disorder, and on admission: the proportion with normal GCS verbal, motor and eye sub-scale scores, the ability to lift both arms against gravity, the ability to walk without the help of another person and urinary incontinence. The

proportion of patients with different pathological types of stroke also varied highly significantly between hospitals. However, these differences should be interpreted cautiously as they are likely to be biased by the large differences in the use of CT scanning between hospitals (refer forward to Table 8.9).

There were no significant differences between cohorts in terms of sex, prior history of diabetes mellitus, ischaemic heart disease or myocardial infarction, independence in activities of daily living prior to the stroke, whether the patient lived alone prior to the stroke or the proportion with a systolic BP of greater than 160 mmHg on admission (although the differences in the latter three variables approached statistical significance: $p = 0.07$, 0.02 and 0.02 respectively).

The prevalence of adverse prognostic variables was highest at Hospital A, lowest at Hospital D and intermediate and similar at Hospitals B, C and E. In particular, of the variables included in the study models, Hospital A had the highest or equal highest proportion of patients with adverse findings (except for whether the patient lived alone prior to the stroke) while Hospital D had the lowest or equal lowest proportion of patients with adverse findings. A substantially greater proportion of patients at Hospital B came from the two most deprived social strata than at any other hospital (73% versus 1 - 29%).

The differences between hospitals in terms of the variables included in the study models is summarised in the different proportions *predicted* to be dead, dead or dependent or alive & at home at six months (Table 7.7). In each case, the predicted

prognosis of the cohort at Hospital D was considerably better than that at any other hospital. The differences in predicted prognosis between the remaining hospitals were less marked. The cohort at Hospital E had the next best predicted prognosis, followed by that at Hospital B, then Hospital C and, worst of all, Hospital A. The absolute differences in the proportion predicted dead, dead or dependent and alive & at home between Hospitals A & D were very large: 16%, 20% and 19%, respectively.

7.4 Response to follow up at six months

7.4.3 Case fatality

Measured by linkage to death certification data (provided by ISD 20 months after the last patient entered the study) there were a total of 1007 deaths by six months amongst the 2724 patients with acute stroke (37% case fatality).

During the data collection phase of the SOP we used a 'hot-pursuit' method to identify deaths to guide our system of follow up (see section 3.8.2). The success of this method (using the linked survival data as a gold standard) is shown in Figure 7.2. The majority of deaths within six months (77%; 777 deaths) occurred in hospital and were identified using SMR1 data. We attempted to identify the 230 out of hospital deaths using our standard method of follow up in 154 cases (see section 3.8) and an enhanced method in 73 cases (see section 6.2); we identified three deaths after discharge using miscellaneous methods. The two methods of follow up identified 80% and 99% of out of hospital deaths, respectively. Of the 1717 cases alive at six

months according to the linked survival data (Figure 7.3), our hot pursuit methods disagreed in only 19 cases (10 reported dead by SMR1, 9 reported dead by GP).

7.4.4 Functional status and place of residence

The methods used to follow up the 1717 patients alive at six months are shown in Table 7.8. The great majority (95%) were contacted using either our standard method (1120 cases) or our enhanced method (508 cases in the follow up trial). The remainder were contacted using one of two other methods: firstly, after the follow up trial had finished, we continued to send a reminder to non-responders (57 cases); and secondly, some patients were simultaneously entered into the SOP and the Feed or Ordinary Diet (FOOD) trial, a randomised trial investigating feeding strategies after acute stroke administered by our department (14 cases) (Dennis 2001). We followed these patients using the FOOD trial methods. We failed to send any follow up in 19 cases (10 because the SMR1 indicated that they had died in hospital, 8 because of administrative error and one because of no GP data).

In total, we collected follow up data from 1257 (73%) of the patients alive at six months (Table 7.8). The overall response to a single follow up was 63% (1075 responses). Examining only the two phases of follow up in which a reminder was sent (the follow up trial and the post follow up trial period) the combined cumulative response after a second follow up was 82% (464 responses from 565 mailings). Response after a single follow up, after a second follow up and after special follow

up is summarised in Figure 7.4. These levels of response are in keeping with those described in chapter six.

Of the 1257 completed outcome questionnaires, 1216 (97%) provided interpretable responses to the questions about dependency in ADL and 1222 (97%) to the questions about place of residence (see Tables 7.9a and 7.9b). The total *useable* response to follow up from patients alive at six months was therefore 71%. Table 7.9a shows that, in returned questionnaires, the principal reason that we failed to collect dependency data was that the simple dependency question was not answered. This contrasts with the residence question (Table 7.9b) in which only eight responses were left blank or were uninterpretable. However, 150 patients had difficulty in selecting a single response to the residency question, although only eight of these multiple responses were uninterpretable.

The median interval between admission to hospital and the completion of an outcome questionnaire for all 1257 responders was 179 days (range between 5^{th} and 95^{th} percentile = 167 to 251 days) and did not vary significantly between hospitals (Table 7.10). The median interval between admission to hospital and the completion of an outcome questionnaire was 202 days and 237 days in the patients in the follow up trial who replied after our second and special contacts, respectively. Only eight questionnaires were completed and returned over one year after the patient had been admitted to hospital. In the 1191 patients in whom we collected follow up data without the use of special methods, 44% of questionnaires were filled by the patient,

52% by a proxy (Table 7.11). As expected, the patient rather than a proxy was more likely to complete the questionnaire if the patient was independent (OR 7.9; 95% CI 5.8 to 10.8) or lived at home (OR 11.6; 95% CI 6.3 to 21.8).

The proportion of patients who were alive at six months and responded to follow up varied highly significantly between hospitals (Table 7.12). Response was greatest at Hospitals A, D and E (79 to 80%), intermediate at Hospital C (71%) and least at Hospital B (63%). Non-responders (n = 460) were significantly more likely than responders (n = 1257) to come from the two most deprived social strata, to be unemployed, and on admission, to have characteristics indicative of severe stroke (Table 7.13).

Table 7.14 shows the comparison of the proportions *predicted* to experience each outcome between responders, non-responders and all patients (i.e. the responders + the non-responders) alive at six months. For each outcome, the proportion predicted to experience a poor outcome was always higher amongst the non-responders than amongst the responders. However, in absolute terms, the predicted risk of the responders only moderately and non-significantly underestimated the predicted risk of each hospital cohort (by up to 5% at Hospital A, up to 4% at Hospital B and up to 2% at Hospitals C, D & E).

7.5 Unadjusted outcomes

Case fatality is reported for all 2724 patients with acute stroke. However, because of losses to follow up, the total number of patients in whom 'death or dependency' can be reported is 2223 (82%: 1007 deaths + 1216 with useable dependency data) and in whom 'alive & at home' can be reported is 2229 (82%: 1007 deaths + 1222 with useable residency data).

There were large and highly significant differences between hospitals in each outcome (shown as proportions in Tables 7.15 (a. to c.) and as unadjusted W scores and odds ratios in Tables 7.16 (a. to c.). Hospital D had, by far, the 'best' set of unadjusted outcomes: the lowest proportions dead and dead or dependent and the highest proportion alive & at home at six months. Hospital A had the 'worst' set of unadjusted outcomes: the highest proportion dead, the second (almost equal) highest proportion dead or dependent and the lowest proportion alive & at home at six months. In absolute terms, the differences in unadjusted outcomes between Hospitals A & D were large: per 100 admissions, 21 more patients were dead, 23 more were dead or dependent and 18 fewer were alive and at home at six months at Hospital A than at Hospital D.

The unadjusted outcomes at Hospitals B,C & E were less consistent. Case fatality was similar at all three hospitals and mid-way between that at Hospitals A & D: about 11 more patients were dead per 100 admitted to Hospital A than to Hospitals B,C & E; and about 10 more patients were dead per 100 admitted to Hospitals B,C &

E than to Hospital D. The proportions dead or dependent at Hospitals A,B & C were almost identical while that at Hospital E was moderately lower (between 4 to 6 fewer patients were dead or dependent per 100 admitted than at Hospitals A,B & C); all were very much higher than at Hospital D. The pattern of outcome for the proportion alive & at home followed that for death or dependency: about 9 more patients per 100 admitted were alive & at home at Hospital D than at Hospital E; and between 4 and 8 more patients per 100 admitted were alive & at home at Hospital E than at Hospitals A,B & C.

7.6 Outcomes adjusted for casemix

7.6.3 The calibration of the study models and the use of W or Ws score methods

The calibration plots of the three study models set are shown in Figure 7.5. The calibration plots of the customised model of case fatality at six months and of the original model for alive & at home at six months follow the 45° line closely, indicating that they are well calibrated. I have therefore reported these outcomes, after adjustment, using ordinary W scores (see section 3.9.3) .

The calibration plot of the model of death or dependency at six months diverges markedly from the 45° line, particularly in the categories with medium and low predicted risks (the model considerably under-predicts the number of patients dead or dependent in patients with a lower than 70% predicted risk). I have therefore

reported adjusted death or dependency using *standardised* W scores (Ws scores; see section 3.9.3).

7.6.4 Adjustment for casemix using the study models

Each adjusted outcome is given in Tables 7.16 (a. to c.). The comparisons of W or Ws scores between hospitals are illustrated in Figures 7.6 (a. to c.). To facilitate these comparisons I have artificially set the W (Ws) score of Hospital A to zero and subtracted the score of Hospital A from the score of each of the other hospitals in each figure. Thus, Figures 7.6 (a. to c.) illustrate the *difference* in W (Ws) score between each hospital and Hospital A. Note, the comparison of W scores indicates the true differences between hospitals while the comparison of Ws scores indicates the differences between hospitals that would be seen *if each had the same casemix structure as the SOP overall.*

For case fatality, after adjustment, the variation between hospitals remained just significant (p 0.047). In absolute terms, the differences in case fatality between Hospital D and Hospitals B,C & E were virtually abolished and the differences in case fatality between Hospital A and the other four hospitals reduced from 21 to between 5 and 7 more deaths per 100 patients admitted.

For death or dependency, after adjustment, the variation between hospitals remained highly significant ($p < 0.0005$) and the ranking of hospitals was unchanged i.e. adjusted death or dependency was significantly lower at Hospital D than at any other

hospital; adjusted death or dependency at Hospital E was significantly lower than that at Hospital B (OR 0.54 (95% CI 0.35-0.85)) but only slightly and non-significantly lower than that at Hospitals A & C. However, *in absolute terms*, the variation in death and dependency between Hospital D and the other four hospitals was halved: the difference between Hospital D and A reduced from 23 to 11 fewer cases dead or dependent per 100 admitted and the difference between Hospital D and E reduced from 18 to 8 fewer cases dead or dependent per 100 admitted.

For alive & at home, after adjustment, the large differences between hospitals reduced very considerably (down to between 2 and 4 more cases alive & at home per 100 admitted) and were no longer statistically significant (p 0.574).

Analyses in which missing predictor variables were replaced using optimistic values; in which cases where stroke occurred *after* admission were excluded; and in which cases with conflicting survival data (record linkage vs. hot pursuit) were re-coded gave almost identical results as our primary analyses (Table 7.17). The exclusion of cases in which our audit assistant had noted particular difficulty in extracting data from the medical record (Table 7.17) did not alter our findings in terms of death or dependency and alive & at home. It did lead the variation in adjusted case fatality between hospitals to become non significant (from p 0.047 to 0.075) but the point estimates of the adjusted odds ratios were almost unaltered.

7.6.5 Alternative adjustments for casemix

The results of the alternative adjustments for casemix are also given in Tables 7.16 (a. to c.). Their comparison with adjustment using the study models is shown in Figures 7.6 (a. to c.) and in Figures 7.7 (a. to c.). Again, the figures illustrate the *difference* in W score or Ws score between each hospital and Hospital A.

These data show that, after adjusting for age, sex and social deprivation, the differences in each outcome between hospitals remained substantial and highly significant ($p \leq 0.0001$ in each case). Similarly, adjustment for urinary incontinence alone explained a much smaller proportion of the variation in each outcome between hospitals than adjustment using the study models. When added to the study models urinary incontinence remained an independent predictor of each outcome ($p < 0.00005$ for its Wald statistic) but did not alter the residual variation in outcome between hospitals.

When added to the study models social deprivation was not an independent predictor of case fatality or being alive & at home (p for its Wald statistic 0.68 and 0.52, respectively). The slight reduction in variation in adjusted case fatality between hospitals that resulted is therefore of dubious significance. However, when added to the study model for death or dependency social deprivation was an independent predictor with greater social deprivation being clearly associated with a greater risk of being dead or dependent (this relationship appeared linear (Figure 7.8) and, for this analysis, I entered social deprivation as a *continuous* variable; $p < 0.0001$ for its

Wald statistic). The additional adjustment for social deprivation led to a modest absolute reduction in the variation in death or dependency between hospitals. As a result, the difference in death or dependency between Hospitals E & B became non-significant (OR 0.74 (0.47 - 1.15).

The adjustment for all measured casemix took account of all the characteristics listed in Table 7.6 except stroke pathological subtype (because of missing values for certain casemix variables this analysis was restricted to 2543 patients for case fatality, 2073 for death or dependency and 2076 for alive & at home). Compared with adjustment using the study models, this adjustment did not alter the absolute differences between hospitals in terms of case fatality or being alive & at home. For death or dependency, it led to a similar reduction in variation as did the addition of social deprivation to the study model i.e. other than social deprivation, none of the remaining casemix variables explained the residual variation in death or dependency.

I omitted stroke pathological subtype (haemorrhagic stroke vs. not) from the 'all casemix' model because of the potential for biased measurement (see Table 8.9). Tested separately, however, the addition of stroke pathological subtype did not account for the residual variation in any outcome (Table 7.18). Its addition did (non-significantly) reduce the variation in death or dependency between Hospital A and Hospitals B & C, but haemorrhage actually appeared to *reduce* the risk of a poor outcome, almost certainly a spurious finding (Table 7.18).

7.6.6 Exploring the differences in adjusted outcome between hospitals

By predicted stroke severity

For case fatality, after adjusting for casemix, there was a suggestion that the worse overall outcome at Hospital A was due to the outcome of patients with severe and, less so, moderate stroke (Figure 7.9a).

For death or dependency, after adjusting for casemix, it was clear that the overall better outcome at Hospital D was due to the outcome of patients with mild and, much less so, moderate stroke (Figure 7.9b).

By the timing of differences in adjusted case fatality

The adjusted survival curves of the five hospitals started to diverge within 5 to 10 days of admission and continued to diverge until about 30 days (Figure 7.10). The curves then ran parallel till about 130 days. At six months, the adjusted survival curves of Hospitals B to E were almost superimposed and clearly separate from the adjusted survival curve of Hospital A.

7.7 Discussion

Prior to adjusting for casemix, there were very large differences between hospitals in the proportion of patients dead, dead or dependent and alive & at home at six months. For each outcome, the results at Hospital D appeared to be considerably 'better' than the results at the other four hospitals. The differences in outcome between Hospitals A,B,C & E were generally less marked, followed a less clear cut pattern and not all were statistically significant. Overall, however, Hospital A tended to have the 'worst' set of outcomes and Hospital E the second 'best'; the results at Hospitals B and C were intermediate to those of Hospitals A and E.

Perhaps not unexpectedly, the large differences in outcome between hospitals were associated with large differences in the baseline characteristics of the patients admitted. The cohort at Hospital D had, by far, the most advantageous set of baseline characteristics while the cohort at Hospital A had the least; the baseline characteristics of the cohorts at Hospitals B, C and E were generally similar and intermediate. Once these differences in casemix were taken into account (using the study models) the overall variation in outcome between hospitals was considerably reduced and, between certain pairs of hospitals, virtually abolished. Much of the variation that remained could be attributed to the play of chance. Indeed, after adjusting for casemix, there were no significant differences in the proportion of patients alive & at home at six months between *any* of the hospitals.

These findings indicate that most of the variation between hospitals in each outcome was due to factors outside of their control rather than to any differences in the quality of the care they provided. Importantly, adjustment only for age, sex and social deprivation, the method currently used to adjust stroke outcomes data published in Scotland, accounted for substantially less of the variation in outcome between hospitals than did our study models; indeed, adjustment for age, sex and social deprivation appeared to be only modestly better than no adjustment at all. These findings strongly suggest that stroke outcomes data as currently published in Scotland – and by extrapolation that in England & Wales - are inadequately adjusted for casemix and hence are unlikely to be accurate indicators of the quality of stroke care.

Nonetheless, despite adjustment using the study models, case fatality and death or dependency at six months continued to vary significantly between hospitals. This residual variation was principally due to Hospitals A and D: case fatality remained significantly higher at Hospital A than at any other hospital (whose case fatalities were almost identical); and death or dependency remained significantly and substantially lower at Hospital D than at any other hospital (whose proportions dead or dependent were similar, except between Hospitals B & E). Given that we used robust prognostic models to adjust for casemix, it is tempting to infer that these residual differences in outcome must be the result of differences in the quality of care. However, as with all observational studies, but especially those based upon routinely and retrospectively derived data, it is likely that other factors may have

been at work. Before drawing any conclusions, therefore, I will first reconsider the quality of our data.

Hospital samples

The first consideration is the validity of our hospital samples. Although case-note retrieval bias seems very unlikely, the possibility that diagnosis and coding bias (see section 1.6.1) may have prevented comparisons of outcome between truly representative samples of patients merits exploration.

Diagnosis bias is possible because we identified cases of stroke simply by accepting the diagnosis recorded in the medical record (see section 3.5.1). False positive diagnosis is more likely if the patient has altered consciousness and false negative diagnosis if the stroke is minor (Ferro *et al.* 1996; Ferro *et al.* 1998b; Libman *et al.* 1995). Such errors may be more likely if the assessing physician is inexperienced (Ferro *et al.* 1998b; Norris and Hachinski 1982) and, to a lesser extent, when there is limited access to CT imaging (Britton *et al.* 1984). These factors may have been more prominent at Hospital A (see chapter 8) and it is possible that its higher case fatality might reflect a greater tendency for its physicians to diagnose moribund non-stroke patients as stroke, and to label patients with minor stroke as a TIA, errors which would spuriously worsen the prognosis of the cohort entered into the SOP. However, even if present, it is unlikely that this bias was large because a bedside diagnosis of stroke is correct in the great majority of cases (Ferro *et al.* 1998b; von

Arbin *et al.* 1981b) and most patients at Hospital A would have had their diagnosis considered by a senior physician shortly after admission.

False positive coding bias cannot influence our findings because we excluded cases who had not in fact had an acute stroke (who we identified reliably – see section 5.3.2). In doing so, it is important to note that we have tested the validity of a system that identifies cohorts of patients with acute stroke more accurately than is currently possible using subsets of routine CVD discharge codes; which is to say, we have tested a system that approximates an improvement in identification of acute stroke likely to result from the routine collection of our casemix variables (see section 5.4). However, our data may still be biased by false negative coding error. Thus, it is arguable that the smaller proportion of patients dead or dependent at Hospital D may have resulted from a greater tendency to assign co-morbid conditions and complications as the primary SMR1 diagnosis in patients with severe stroke (spuriously *improving* the prognosis of the cohort entered into the SOP) and, similarly, that the higher proportion of patients dead at Hospital A may have resulted from the coding of minor stroke as TIA (spuriously *worsening* the prognosis of the cohort). Again, however, even if present, it seems unlikely that this bias can explain much of the residual variation in outcome: a previous study suggests that false negative coding affects only about 14% of acute strokes at Hospital D, that only a third of these have a stroke code listed as a secondary diagnosis and that their exclusion does not bias the estimates of case fatality or discharge home (Davenport *et al.* 1996c); and our own audit revealed that a primary SMR1 diagnosis of TIA had

a very low positive predictive value for stroke i.e. strokes were not being 'hidden' within that code at any hospital.

Casemix data

The second consideration is the possibility that the validity of our data describing casemix, and hence of our adjustments for casemix, may have varied between hospitals. In many respects, the quality of our casemix data was good: we were able to abstract a complete set in virtually all cases; the replacement of the small number of missing values was not a source of bias; the data pooled from all five hospitals were reliable (see section 5.3.2); and the data abstracted at Hospital D were valid (see section 5.3.3). However, I did not directly address the question of variation in the validity of the data between hospitals and across patient types. The possibility remains, therefore, that variation in the methods used by clinicians to collect and record baseline characteristics and by our research assistant to extract those data may have been a source of bias. Furthermore, there is reason to suspect that any such error may have been most marked at Hospital A, and hence that its higher adjusted case fatality may in fact reflect our failure to properly adjust for its 'sicker' patients.

First, Hospital A admitted the greatest proportion of patients with severe stroke, many of whom died within a few days of admission. Clinical data collection has been shown to be less accurate in such patients (Shinar *et al.* 1985) and it may also be less comprehensive. Second, the accuracy and completeness of data recorded in the medical record is influenced by the training and knowledge of the clinician

(Rowley and Fielding 1991b; Schmulling *et al.* 1998) and by the use of specific prompts (such as those found on a stroke clerking proforma) (Davenport *et al.* 1995b). Hospital A differed from the other hospitals in both these regards (refer forward to Table 8.1). Third, our audit assistant may have been influenced by the knowledge of in-hospital death when abstracting data from the medical record; a bias which, if present, would be most marked at Hospital A; and lastly, our audit assistant noted greater difficulty in extracting data from the medical record (and so perhaps needed to make more inferences and extrapolations in order to extract predictive data) at Hospital A than at any other hospital. Indeed, after the exclusion of the cases which were difficult to audit, the residual differences in case fatality between hospitals were no longer significant.

Two points can be made against this argument. First, the exclusion of cases which were difficult to audit did not alter the point estimates of the odds ratios of adjusted case fatality, suggesting that their exclusion may only have reduced the certainty of the original findings rather than demonstrated the presence of bias; and second, where the medical record made it clear that the patient had had a severe stroke or had died, one might imagine that our audit assistant would have been most likely to make overly pessimistic assessments of casemix (Caplan *et al.* 1991b; Gjorup *et al.* 1986), leading us to over rather than under adjust. Nonetheless, the possibility of bias in our adjustments for casemix, particularly at Hospital A, remains.

Outcome data

The next consideration is the validity of our measurements of outcome. Given its unambiguous nature and collection by record linkage, our measurements of survival are likely to be accurate. However, our measurements of place of residence and functional status are open to bias because they depend on the measurement properties of the measuring instruments and on the response to our outcome questionnaire.

We measured dependency in ADL using the simple dependency question. Although a validated instrument it is only moderately precise (see section 3.7.2), and, as already discussed, its use may have led us to miss moderate differences in death or dependency (see section 6.4). The use of the simple dependency question may therefore have contributed to the limited variation in adjusted death or dependency between Hospitals A,B,C & E; however, the markedly lower adjusted death or dependency at Hospital D cannot be explained by a second order bias such as this.

Our method of measuring place of residence (a menu of possible responses) has not been validated and led 12% of survivors to give multiple responses. Fortunately the simple dichotomisation of place of residence into institutional care versus all other forms of accommodation ('home') allowed sensible interpretation of virtually all multiple responses. However, dichotomisation also reduced the degree to which residence data were able to approximate level of disability. This may partly explain the failure of the alive & at home data to mirror the death or dependency data, after adjustment for casemix. In particular, it is possible that the dichotomisation may

have hidden a superior residential outcome at Hospital D, where significantly fewer survivors were dependent at six months. Thus, institutionalised patients at Hospital D may have primarily lived in residential homes whilst those elsewhere may have made greater use of nursing homes or long term hospital care; and non-institutionalised patients at Hospital D may have primarily lived in their original home whilst those elsewhere may have been more likely to move, for example, to live with family or in sheltered accommodation.

Given the high and variable level of non-response across hospitals to our survey of outcome in survivors, the possibility of non-response bias is clear. Certainly, as a group, non-responders had a higher prevalence of adverse prognostic variables (see Table 7.13) and, in the follow up trial, had non-significantly worse outcomes than responders (see section 6.3.3). More importantly, even after two follow ups, the follow up trial also showed that there may be important bias in comparisons of dependency between some hospitals (see section 6.3.3). As such, the potential for non-response bias would seem to be greatest at Hospitals B & C, the two hospitals with the lowest levels of response (63% and 71%, respectively).

Against this, however, the follow up study also showed that, even at low levels of response, many estimates of dependency and all estimates of institutionalisation were only modestly biased (see section 6.3.3). Also, in the analysis in Table 7.14, the differences in the proportion *predicted* 'dead or dependent' and the proportion *predicted* 'alive & at home' between the responders and the entire population at

each hospital and overall were moderate (0 to 5% in absolute terms) and none were significant. Our main analyses are also adjusted for casemix, further reducing the impact of non-response (see section 6.4). Hence, whilst the possibility of substantial non-response bias cannot be excluded, it seems most likely that it was moderate. Specifically, whilst non-response may explain some of the variation, or lack of variation, in adjusted death or dependency between Hospitals A,B,C & E it seems an implausible explanation for the very much lower adjusted death or dependency at Hospital D.

Adjustment for casemix

The final potential source of error is failure to properly adjust for differences in casemix. Assuming that all the predictor variables are unbiased, the ability to adjust for casemix depends on the statistical and clinical validity of the predictive model (Altman and Royston 2000). A statistically validated model is 'one which passes all appropriate statistical checks, including goodness of fit on the original data set and unbiased prediction on a new data set' (Altman and Royston 2000). By this definition, our models all showed good statistical validity prior to the start of the study (see section 3.9.1). The models predicting case fatality and the proportion alive & at home also calibrated very well within the SOP, providing further evidence of their external validity and allowing unbiased comparisons of these outcomes using ordinary W scores. The model predicting case fatality can be criticised because it was customised within the SOP in which about 6% of patients had two or more admissions i.e. the model transgressed the assumption that all predictor variables in a

logistic regression analysis are independent of each other. However, given the small numbers involved it is likely that any resulting error is small.

The model predicting death or dependency calibrated poorly within the SOP, under-estimating the risk of death or dependency in all groups but especially in those with a truly low or intermediate risk. The explanation for this is uncertain and may relate to deficiencies in the validity, reliability and completeness of our data. However, it is notable that the model also gave 'over-optimistic' predictions of death or dependency in groups at low or intermediate risk in the community cohort first used to establish its external validity (Counsell *et al.* 2002). Together, these observations suggest that additional predictive data may be needed for the model to accurately predict death or dependency in lower risk groups. From a practical perspective, ordinary W scores calculated using the model would have provided a biased comparison of adjusted death or dependency, particularly disfavouring Hospital D (the hospital which admitted the greatest proportion of patients at low or intermediate baseline risk). Fortunately, this was prevented by the use of the standardised W score and so the *comparison* of adjusted death or dependency between hospitals is valid.

Clinical validation refers to whether or not a model 'performs satisfactorily on a new data set according to context-dependent statistical criteria laid down for it' (Altman and Royston 2000). In the SOP, the task was to provide sufficient adjustment such that the majority of any residual variation in outcome might confidently be ascribed to differences in the processes of stroke care. The problem, of course, is that in the

absence of perfect models with which to compare the performance of our own, it is impossible to know how close they come to fulfilling this aim. Three factors suggest that our models may have good clinical validity. First, the models explained large amounts of the variation in each outcome i.e. the predictive data intrinsic to each model appears to be strong. Second, after adjustment, the residual differences in outcome between hospitals were of the same order of magnitude as that found in a systematic review of randomised trials comparing stroke unit care with care on general wards i.e. it is plausible that differences in stroke care can produce differences in outcome of roughly these sizes (Mant and Hicks 1996; Stroke Unit Trialists' Collaboration 1997a). Third, only a small amount of the residual variation in outcome could be explained by further adjustment for other measured and potentially important markers of casemix. The failure of urinary incontinence to add to the predictive power of our models and the clear inferiority of adjusting for urinary incontinence alone is particularly interesting given its recognised importance as a powerful predictor of outcome after stroke; it also provides somewhat indirect evidence that our models may be superior to some previous models that predict outcome after stroke (Barer and Mitchell 1989; Gladman *et al.* 1992a; Taub *et al.* 1994).

The only instance when the addition of another variable resulted in a material (although still modest) improvement in adjustment was when social deprivation was added to the model predicting death or dependency. Its addition suggested that the apparently worse 'performance' of Hospital B as compared to Hospital E was an artefact of the populations served. The improvement in the predictive power of the

model is most likely to have resulted from an improvement in the prediction of dependency since social deprivation was not an independent predictor of case fatality. The improvement is credible if one considers that dependency reflects the interplay of the patient's disability and their social environment (Wilkin 1987). Other investigators (Jakovljevic *et al.* 2001; van den Bos *et al.* 2002) (but not all (Fullerton *et al.* 1988)) have found a similar relationship and a previous study that tested the addition of social deprivation to our model also found a non-significant trend toward it being an independent predictor of death or dependency (Counsell 1998) (the failure to reach significance perhaps reflecting the small proportion (8%) of patients in the two lowest socio-economic groups in this study). The addition of social deprivation to our model therefore appears worthy of further investigation.

Clearly, our models fail to take account of many other factors relating to the patient's social background, psychological status, past medical and stroke history, clinical examination, laboratory investigations and brain imaging that have at one time or another been put forward as independent predictors of outcome after stroke and which were not measured in the SOP. The importance of these omissions is difficult to gauge. Many were considered during the development of our models and were not found to be independently associated with outcome, or, even if they were, they did not add to the models' predictive power (Counsell *et al.* 2002; Counsell 1998). Also, few were found to be consistent predictors of outcome in a recent systematic review of multivariate prognostic models (Counsell and Dennis 2001). On the other hand, the developers of our models were clearly unable to consider all known predictive factors and the frequent poor methodological quality of the studies in the systematic

review allowed for few definitive conclusions (Counsell and Dennis 2001). Hence, whilst our models are undoubtedly very good, not to say state of the art, caution remains sensible and one must be careful not to succumb to 'the fallacy of omnimetrics', namely "the ill-conceived idea that if one just identifies the right things to measure and develops the right scales of measurement, then one can determine all that one needs to know about prognosis and can make valid treatment comparisons" (Green and Byar 1984). It cannot be overemphasised that, in the context of comparisons between non-randomly allocated groups, one can never discount the possibility, indeed probability, that a proportion of the residual variation in outcome may be due to variation in factors that have not been considered or which remain to be understood. In support of this possibility, it is notable that our methods did not pick out Hospitals A and D from the generality of hospitals. Rather, these hospitals had the most extreme outcomes *before* adjustment for casemix. The nagging doubt remains, therefore, that their continued status as the hospitals with the most extreme outcomes *after* adjustment may simply reflect the difficulties of adjusting for all variations in casemix when those variations are very large.

The play of chance

The final consideration is that the residual differences in outcome may be due to the play of chance. Clearly, given the p value of < 0.0005, this is an extremely unlikely explanation for the residual differences in death or dependency. However, the residual differences in case fatality were only just significant (p 0.047), a result so close to the traditional cut-off of $p < 0.05$ that it must remain plausible that Hospital A appeared to have the highest adjusted case fatality simply because of bad luck

(Sterne and Davey Smith 2001). Indeed, the use of the traditional $p < 0.05$ level of statistical significance might be criticised because the study measures three outcomes and a shrunken p value (e.g. $p < 0.017$, generated using the Bonferroni correction) should perhaps have been used to minimise the risk of a Type 1 error (Anonymous 1996;Anonymous 1999). Similarly, the use of binomial methods to calculate the confidence intervals for our W scores fails to account for the uncertainty in the adjusted outcome that results from the uncertainty in the estimated co-efficients of our prognostic models (Hosmer and Lemeshow 1995; Signorini and Weir 1999) i.e. it is arguable that the confidence intervals of our W scores should overlap somewhat more than appears to be the case. If we had used these stricter methods, the residual difference in case fatality between Hospital A and the other four hospitals would certainly not have been significant. It is debatable, however, whether such strict statistical methods are appropriate to the field of quality assessment where the aim is to identify hospitals likely to have outcomes significantly different to the remainder, not to prove it beyond all reasonable doubt. Also, we collected data only over a two rather than three year period (as is the case for the stroke outcomes data published in Scotland). It is possible that the finding of lower adjusted case fatality at Hospital A would have been more certain if we had collected data for another year.

Conclusion

Having considered our methods in some detail, it is clear that no aspect of our comparison of outcomes escapes criticism. This is, of course, not a surprise given that the SOP is an observational study quite purposefully based on routinely and retrospectively derived data and specifically limited in the detail and tenacity of its

measurements of outcome. These factors have undoubtedly led to a considerable number of potential biases which differ in their extent and direction across the study hospitals. Although none are large in themselves, the impact of their accumulation is difficult to predict and, in combination, it is quite possible that they may explain much of the residual variation in outcome between hospitals. As is always the case, variation in unmeasured baseline characteristics may also explain some of the remaining variation in outcome and the play of chance is a plausible explanation of the residual variation in case fatality. To return to the original question, therefore, are these data sufficiently robust that one can conclude that the residual differences in outcome between the study hospitals *are* the result of differences in the quality of care? Clearly, the answer must be no. The fact that so many misgivings remain despite our efforts to use an idealised system of measuring outcome underlines the enormous difficulties inherent to this approach of measuring the quality of stroke care.

Nonetheless, as previously noted, it is also true that the purpose of the measuring outcomes is to indicate hospitals likely to have especially poor or excellent care, not to prove that this is the case beyond all reasonable doubt. By demonstrating that there are no biases that *definitively* explain our findings, this review leaves open the possibility that much of the residual variation in outcome might be due to variation in the quality of stroke care. The divergence of the adjusted hospital survival curves over the period from 5 to 10 days until 30 days lends some weight to this possibility – this is the period after a stroke when death is most likely to result from the complications of immobility and cardiovascular disease, disorders potentially

preventable by appropriate intervention (Bamford *et al.* 1990b; Davenport *et al.* 1996a; Stroke Unit Trialists' Collaboration 1997b). However, whether the residual differences in adjusted outcome are truly associated with differences in the quality of stroke care can only be determined by its direct inspection. If adjusted outcomes really are indicators of the quality of stroke care, one would expect to find deficiencies in aspects of care that promote survival at Hospital A; excellence in aspects of care that promote independence in ADL at Hospital D; and broad similarities in care at Hospitals B,C & E. These hypotheses are considered in the chapter that follows.

Summary

Major findings

1. Prior to adjustment for casemix there were very large differences between hospitals in the proportion dead, dead or dependent or alive & at home at six months. These differences were associated with large differences between hospitals in the measured baseline characteristics of the patients admitted.

2. Once these differences in casemix were taken into account, the variation in each outcome reduced considerably and much that remained could be attributed to the play of chance. After adjustment, there were no significant differences in the proportion alive & at home between any hospitals. Variation in outcomes after stroke appears to be mostly due to factors that are outside of hospital control.

3. After adjustment, compared with the other hospitals, case fatality remained significantly higher at Hospital D and the proportion dead or dependent remained significantly and substantially lower at Hospital A. These differences might reflect differences in the quality of care but equally might simply reflect biased measurements, inadequate adjustment for casemix and, for case fatality, the play of chance.

4. If the residual differences in outcome are the result of differences in quality of care, one would expect to find: deficiencies in aspects of care that promote

survival at Hospital A; excellence in aspects of care that promote independent survival at Hospital D; and broad similarities in care at Hospitals B, C & E.

Other findings

1. This study provides further evidence of the external validity of our models for the prediction of case fatality and being alive & at home at six months. It also confirms a previous observation that the model predicting death or dependency at six months may be inaccurate ('over-optimistic') in patients at low and intermediate risk.

2. The predictive power of the study models appears to be considerably greater than that of urinary incontinence alone and is not improved by its addition. Social deprivation appeared to be an independent predictor of death or dependency when added to the study model and modestly improved its predictive power. These findings require further study in a large, prospectively collected data set.

3. The methods used to adjust the stroke outcomes data currently published in Scotland and in England & Wales are clearly inadequate. Any inferences relating to the quality of stroke care based upon them are likely to be misleading.

4. Data describing the six predictive variables included in the study models, urinary incontinence and social deprivation can be abstracted from the medical records of virtually all patients with acute stroke.

5. In the absence of a link to central death certification data, a 'hot-pursuit' contact with the GP is likely to identify about 80% of deaths that occur after discharge from hospital. The addition of a rapid link to centrally held death certification data is likely to increase this to about 99%.

6. The proportion of questionnaires in which the simple dependency question is left blank or mis-filled is very low. However, 12% of respondents found the menu of responses for the residence question confusing; a simple question asking whether the patient lives in a hospital, nursing or residential home might be preferable.

Table 7.1 Hospital stays with an ICD cerebrovascular disease code given as the primary diagnosis and their audit at each hospital

Assignment to hospital based on SMR1 data

	OVERALL		Hospital A		Hospital B		Hospital C		Hospital D		Hospital E	
	n	%	n	%	n	%	n	%	n	%	n	%
Hospital stays identified	4223		740		1187		717		999		580	
Audited	4017	**95**	696	**94**	1114	**94**	700	**98**	967	**97**	540	**93**
Not audited												
1. Did not inspect the medical record	182		42		65		14		21		40	
2. Insufficient data within medical record *	24		2		8		3		11		0	

* Insufficient data to allow a diagnosis of either stroke, ineligible stroke or not stroke to be determined.

Table 7.2 Comparison of characteristics and outcomes of patients with audited and un-audited hospital stays; overall and per hospital

Assignment to hospital, baseline characteristics and type of admission based on SMR1 data

	OVERALL			Hospital A			Hospital B			Hospital C			Hospital D			Hospital E		
	Audit	No audit	p	Audit	No audit	p	Audit	No audit	p	Audit	No audit	p	Audit	No audit	p	Audit	No audit	p
Median age	73	75	**0.08**	75	77	0.29	73	73	0.54	75	77	0.25	69	70	0.87	74	74	0.87
Male %	48	42	**0.07**	48	36	0.16	47	37	0.11	50	47	1.00	49	41	0.42	47	55	0.42
Emergency admissions %	84	77	**0.01**	96	98	0.89	74	63	0.05	86	82	0.91	79	75	0.79	94	78	**0.01**
Dead at six months %	32	32	**1.00**	43	39	0.65	31	36	0.52	34	6	0.03	23	28	0.68	33	33	1.00
Dead or dependent at six months % *	79	81	**0.83**	86	83	0.76	88	92	0.76	87	83	0.67	59	65	0.80	82	77	0.53

p Mann-Whitney U test for comparisons of age; χ^2 test with continuity correction for comparisons of proportions except for 'death or dependency' at individual hospitals where I used Fisher's Exact test.

* Analysis performed using the 2797 patients in whom we collected data describing dependency

Table 7.3 The diagnoses of patients with audited hospital stays; overall and per hospital

Diagnosis	OVERALL		Hospital A		Hospital B		Hospital C		Hospital D		Hospital E	
	n	%	n	%	n	%	n	%	n	%	n	%
1. All stays (hospital by SMR1)*	4017		696		1114		700		967		540	
Not stroke	1012	**25**	155	**22**	328	**29**	142	**20**	244	**25**	143	**26**
Ineligible stroke	160	**4**	15	**2**	25	**2**	24	**3**	86	**9**	10	**2**
Stroke	2845	**71**	526	**76**	761	**68**	534	**77**	637	**66**	387	**72**
2. Stroke (hospital by audit data)*	2845	**71**	532	**76**	762	**68**	533	**77**	623	**64**	395	**73**
3. Non SAH versus SAH stroke												
Non SAH stroke	2724	**96**†	521	**98**†	746	**98**†	520	**98**†	551	**88**†	386	**98**†
SAH stroke	121	**4**†	11	**2**†	16	**2**†	13	**2**†	72	**12**†	9	**2**†

* SMR1 and audit data disagreed about the hospital to which the patient was first admitted in a small number of cases of acute stroke (see table 7.4); where there was disagreement, I assumed that the audit data were correct and altered the patient's hospital assignment (similar alterations for not stroke and ineligible stroke cannot be made because we did not collect data regarding the true admitting hospital in these cases).

† Denominator for calculation of % = all strokes with hospital assignment by audit data.

Table 7.4 Disagreement between SMR1 and medical record audit regarding the hospital to which the patient was first admitted (acute strokes only)

	First hospital by audit of medical record					
First hospital by SMR1	Hospital A	Hospital B	Hospital C	Hospital D	Hospital E	Total
Hospital A	525			1		526
Hospital B		761				761
Hospital C		1	533			534
Hospital D	7			622	8	637
Hospital E					387	387
Total	532	762	533	623	395	2845

Total disagreement on first hospital of admission for acute strokes = 17 cases

(8 SAH, 6 haemorrhagic stroke, 2 ischaemic stroke, 1 pathological type unknown).

Table 7.5 Number of missing values for key baseline characteristics for acute stroke patients per hospital.

Characteristic	Number of cases with missing values					
	Total	Hosp. A	Hosp. B	Hosp. C	Hosp. D	Hosp. E
Study model variables						
Age (on admission)	0	0	0	0	0	0
Independent in ADLs	6	0	1	0	4	1
Lived alone	9	0	2	0	6	1
Normal verbal GCS	15	3	5	0	5	2
Can lift both arms against gravity	12	1	4	0	4	1
Can walk without help of another	7	0	2	0	4	1
One or more study model variable*	**19**	**3**	**5**	**0**	**9**	**2**
Others						
Social class	19	7	1	3	8	0
Urinary incontinence	14	1	4	1	7	2

* Cases with missing values are not mutually exclusive

Table 7.6a Baseline characteristics of the patients with non-SAH acute stroke

Characteristic	Total d	Hospital A		Hospital B		Hospital C		Hospital D		Hospital E		p (χ^2)
		d	%	d	%	d	%	d	%	d	%	
Median age (IQR)	2724	**75**	(67 to 82)	**73**	(65 to 80)	**75**	(67 to 82)	**73**	(63 to 82)	**75**	(67 to 81)	0.002 KW
Sex (Male)	2724	521	**49**	746	**46**	520	**50**	551	**50**	386	**47**	0.61
Social class	2705	514		745		517		543		386		<0.0001
DepCat 1 - 2	266		**7**		**2**		**7**		**29**		**7**	
DepCat 3 - 5	1666		**89**		**25**		**64**		**62**		**92**	
DepCat 6 - 7	773		**4**		**73**		**29**		**9**		**1**	
Prior to the stroke												
Independent in ADLs	2718	521	**84**	745	**88**	520	**84**	547	**88**	385	**88**	0.07
Lived alone	2715	521	**36**	744	**39**	520	**41**	545	**32**	385	**38**	0.02
Employed	2718	521	**7**	745	**4**	520	**5**	547	**14**	385	**8**	<0.0001
Previous history of												
Diabetes mellitus	2717	521	**15**	745	**15**	520	**15**	547	**12**	384	**14**	0.64
Ischaemic heart dis.	2717	521	**34**	745	**37**	520	**38**	547	**33**	384	**34**	0.33
Myocardial infarction	2717	521	**17**	745	**16**	520	**18**	547	**17**	384	**16**	0.95

d = denominator, IQR = inter-quartile range, KW = Kruskal Wallis test

The denominators do not sum to a total of 2724 in cases where there were missing data

Table 7.6 cont.

Characteristic	Total d	Hospital A		Hospital B		Hospital C		Hospital D		Hospital E		p (χ^2)
		d	%	d	%	d	%	d	%	d	%	
Days from stroke onset to admission median (10-90 centiles)	2566 †	**0**	(0 to 2)	**0**	(-0.5 to 2)	**0**	(0 to 2)	**0**	(0 to 4)	**0**	(0 to 2)	<0.0001 KW
Stroke onset in hospital	2724	521	**5**	746	**9**	520	**6**	551	**7**	386	**5**	0.006
Stroke type	2724	521		746		520		551		386		<0.0001
Ischaemic	1754		**46**		**67**		**68**		**74**		**67**	
Haemorrhagic	301		**6**		**10**		**12**		**13**		**15**	
Unknown	669		**48**		**23**		**20**		**13**		**18**	
On admission												
Normal verbal GCS	2709	518	**44**	741	**52**	520	**49**	546	**65**	384	**53**	<0.0001
Normal eye GCS	2708	518	**70**	742	**76**	520	**74**	545	**83**	383	**76**	<0.0001
Normal motor GCS	2708	518	**69**	742	**77**	520	**77**	545	**87**	383	**79**	<0.0001
Can lift both arms against gravity	2712	520	**45**	742	**51**	520	**53**	545	**71**	385	**54**	<0.0001
Can walk without help of another	2717	521	**17**	744	**17**	520	**17**	547	**41**	385	**18**	<0.0001
Sys. BP ≥ 160 mmHg	2687	513	**38**	741	**37**	520	**37**	537	**41**	376	**47**	0.02
Urinary incontinence	2710	520	**62**	742	**63**	519	**62**	544	**47**	385	**65**	<0.0001
Difficulty auditing the medical record	2714	520	**13**	744	**7**	520	**5**	545	**9**	385	**6**	<0.0001

d = denominator; KW = Kruskal Wallis test; Sys. BP = Systolic BP

† d per hospital for this variable) A to E: 487,704,494,524,357; missing cases are those where I could not identify the exact interval from onset to admission but was clear that it was < 30 days.

Table 7.7 Predicted outcomes at six months after admission per hospital

Hospital		Proportion predicted with each outcome at six months					
		Dead		Dead or dependent		Alive & at home	
	n	%	p	%	p	%	p
Hospital A	521	**43**	224.9	**78**	403.8	**41**	211.7
Hospital B	746	**38**	280.9	**74**	548.8	**45**	338.9
Hospital C	520	**40**	205.5	**75**	391.9	**43**	222.9
Hospital D	551	**28**	151.4	**58**	317.8	**60**	328.9
Hospital E	386	**37**	144.4	**72**	279.3	**46**	177.3

p = total number of cases predicted to have an outcome, calculated as the sum of the individual predictions of risk generated for each patient by the study prognostic models.

Table 7.8 Response to the various methods of follow up used in patients with acute stroke who were alive at six months

Method of follow up allocated	Number of patients followed	Cumulative response % (n)		
		After single mailing	**After second mailing**	**After special follow up**
1. Standard	1120	**64** (714)		
2. Follow up trial	508*	**62** (317)	**83** (420)	**96** (486)
3. Post follow up trial	57	**54** (31)	**77** (44)	
4. FOOD trial †	14	93 (13)		
5. Follow up omitted in error	18			
TOTAL	1717	**63** (1075)	**69** (1191)	**73** (1257)

* The follow up trial included a small number of patients who, in hindsight, were ineligible for the SOP (because they were first admitted to a non-study part of Hospital A) who were retained in the analysis to maintain the comparison of truly randomly allocated groups. However, these patients were NOT retained within the main SOP analyses. Hence there is a difference between the number of patients with stroke who were alive at six months in the follow up trial (512) and in the SOP overall (508).

† Patients simultaneously entered into the SOP and the FOOD trial. One patient in the FOOD trial was not followed in error, hence the total number of patients not followed in error is 18 + 1 = 19 cases.

Table 7.9a The proportion of responses to the questions about dependency from which useable outcome data could be obtained

1257	Patients with outcome data collected

To the question: " Do you need help from anybody with everyday activities?"

1201	Replied Yes or No
53	Left the question blank (one due to a printing error)
3	Gave an unintelligible response
56	Unusable responses

Of these 56 patients, 15 correctly completed the Modified Rankin Scale

∴ **Useable response** to dependency questions

= 1201 + 15 = **1216**

= **97%** of patients with outcome data collected (denominator = 1257)
= **71%** of patients followed & who were alive at 6 months (d = 1717)

64 of the 66 patients who responded to special follow up gave dependency data
∴ response (as above) to mailed survey = 1152/1191 = 97%
= 1152/1717 = 67%

Table 7.9b The proportion of responses to the questions about residence from which useable outcome data could be obtained

1258	Patients with outcome data collected
17	No inquiry about residence made *
1240	Response to residence question as below

Response to multiple option residence question (Y = yes)					Number	**Interpretation** †
At Home	With family	Other place	Hospital	Residential care/ Nursing Home		
Y	-	-	-	-	823	Home
Y	Y	-	-	-	135	Home
Y	-	Y	-	-	4	Home
Y	-	-	Y	-	1	?
Y	-	-	-	Y	2	?
Y	Y	-	Y	-	2	?
Y	Y	-	-	Y	1	?
Y	-	Y	Y	-	1	?
-	Y	-	-	-	52	Home
-	Y	Y	-	-	1	Home
-	Y	Y	Y	Y	1	?
-	-	Y‡	-	-	25	Home
-	-	Y‡	-	-	2	?
-	-	Y‡	-	-	2	Not home
-	-	Y	Y	Y	1	Not home
-	-	-	Y	-	70	Not home
-	-	-	Y	Y	1	Not home
-	-	-	-	Y	108	Not home
					8 blank/ confused	?

* Patients in both the SOP and the FOOD trial whose outcome questionnaire did not inquire about residence. We used FOOD trial data to identify the outcome in these cases (13 followed up using FOOD methods only; 4 both FOOD and SOP follow up sent but reply to FOOD questionnaire only).

† Home = 'at home' or 'with family'; 'Not home' = 'hospital' or 'residential/nursing care'. Interpretation of response required in cases of multiple response.

‡ Where the only response was 'other place', I used the other details supplied to define home vs. not.

Useable response = 823 + 135 + 4 + 52 + 1 + 25 + 2 + 1 + 70 + 1 + 108 = **1222**

= **97%** of patients with outcome data collected (denominator = 1257)
= **99%** of patients in whom residence questioned (d = 1240)
= **71%** of patients followed & who were alive at 6 months (d = 1717)

All 66 patients who responded special follow up gave residence data
∴ useable response by mailed survey = 1156 (response 97%, 98%, 67% respectively)

Multiple response to residence question = 150 patients (12%)

Table 7.10 Interval from the date of admission to the date the follow up questionnaire was completed

Six months = 183 days

Method of follow up	Number of respondents	Interval from admission to completion of questionnaire (days)		Difference between hospitals? p (Kruskal Wallis)
		Median	5th – 95th percentiles	
1. Standard	714	179	169 - 276	0.60
2. Follow up trial				
Responders to first mailing	317	172	165 – 188	0.04
Responders to second mailing	103	202	181 – 226	0.71
Responders to special follow up	66	237	229 – 290	0.19
3. Post follow up trial	44	180	171 – 258	0.76
4. FOOD trial	13	185	174 – 288	*
OVERALL	1257	**179**	**167 – 251 ‡**	**0.37**

* Only patients from Hospital D were entered into the FOOD trial during the SOP

‡ Range = 150 to 688 days; 8 cases returned a questionnaire over 1 year after admission (at 367, 376,377,392,400,402,438 & 688 days, respectively).

Table 7.11 Who completed the follow up questionnaire

Of the patients with acute stroke alive at six months who were followed without the use of special methods (n = 1191).

Respondent	**Response** % (n)						
	Overall	**By functional status**			**By residence**		
		Dependent	Independent	No data	At home	Not at home	No data
Patient	**44** (521)	**30** (239)	**74** (262)	51 (20)	**49** (495)	**8** (13)	38 (13)
Proxy	**52** (625)	**67** (536)	**21** (74)	39 (15)	**47** (465)	**89** (142)	53 (18)
Both	**1** (10)	1 (8)	1 (2)	0 (0)	1 (8)	1 (2)	0 (0)
Not specified	**3** (35)	2 (18)	4 (13)	10 (4)	3 (29)	2 (3)	9 (3)

Odds ratio of patient completing the questionnaire if independent vs. dependent = **7.9** (95% CI: 5.8 – 10.8)

Odds ratio of patient completing the questionnaire if at home vs. not at home = **11.6** (95% CI: 6.3 – 21.8)

(Odds ratios of response calculated for each outcome using the appropriate table data where the % response is shown in bold)

Table 7.12 Total and useable response to follow up by patients alive at six months per hospital

	Number followed	Total response % (n)	Useable response % (n)	
			Dependency question	**Residence question**
Hosp. A	271	**79** (215)	**76** (207)	**78** (211)
Hosp. B	473	**63** (297)	**61** (287)	**62** (291)
Hosp. C	325	**71** (232)	**69** (225)	**71** (230)
Hosp. D	404	**79** (318)	**77** (313)	**74** (299)
Hosp. E	244	**80** (195)	**75** (184)	**78** (191)
TOTAL	1717	**73** (1257)	**71** (1216)	**71** (1222)
Comparison of % response between hospitals p (χ^2)		< 0.0001	< 0.0001	< 0.0001

Table 7.13 Comparison of baseline characteristics between responders and non-responders to follow up (for patients with acute stroke alive at six months)

Responder = a patient who responded to any method of follow up (n = 1257)

Characteristic	Responder		Non-responder		Difference
	%	d	%	d	p χ^2
Demographics					
Age median (IQR)	**72**	(62-79)	**71**	(63-79)	0.762 MW
Male	**51**	1257	**50**	460	0.746
Social deprivation (by DepCat)		1247		459	**<0.0005**
1 to 2	**11**		**7**		
3 to 5	**63**		**54**		
6 to 7	**26**		**39**		
Prior to stroke					
Independent in ADLs	**93**	1252	**91**	460	0.093
Lived alone	**36**	1251	**40**	459	0.110
Employed	**12**	1252	**7**	460	**0.005**
Diabetes mellitus	**13**	1252	**15**	460	0.227
Ischaemic Heart Disease	**34**	1252	**32**	460	0.474
Previous myocardial infarction	**17**	1252	**17**	460	0.924
Admitted to		1257		460	**<0.0005**
Hosp. A	**17**		**12**		
Hosp. B	**24**		**38**		
Hosp. C	**18**		**20**		
Hosp. D	**25**		**19**		
Hosp. E	**16**		**11**		
On admission					
Normal GCS Verbal score	**73**	1251	**60**	458	**<0.0005**
Motor score	**96**	1250	**91**	458	**<0.0005**
Eye score	**93**	1250	**88**	458	**0.001**
Able to lift both arms against gravity	**75**	1250	**62**	459	**<0.0005**
Able to walk without assistance	**35**	1252	**25**	459	**<0.0005**
Systolic BP ≥ 160 mmHg	**40**	1245	**40**	458	0.902
Urinary incontinence	**38**	1249	**57**	458	**<0.0005**
Stroke onset in hospital	**2**	1257	**2**	460	0.906
Pathological type of stroke		1257		460	**0.002**
Ischaemic	**81**		**75**		
Haemorrhagic	**11**		**11**		
Unknown	**8**		**14**		

d = denominator (varies due to missing data) IQR = inter-quartile range; MW = Mann Whitney U test

Table 7.14 The comparison of the baseline predicted risks of each outcome between cohorts of responders, non-responders and all patients alive at six months

	Proportion PREDICTED to experience each outcome* %			Responders vs. all cases p (χ^2)
	Responders	Non-responders	All cases	
Hospital A				
Dead	20	40	25	0.31
Dead or dependent	59	77	63	0.40
Alive & at home	62	44	58	0.44
n	215	56	271	
Hospital B				
Dead	22	30	25	0.35
Dead or dependent	59	70	63	0.24
Alive & at home	61	50	57	0.25
n	297	176	473	
Hospital C				
Dead	25	33	27	0.51
Dead or dependent	65	69	66	0.77
Alive & at home	55	50	54	0.70
n	232	93	325	
Hospital D				
Dead	17	21	17	0.73
Dead or dependent	46	50	47	0.82
Alive & at home	71	68	71	0.88
n	318	86	404	
Hospital E				
Dead	23	34	25	0.64
Dead or dependent	61	67	62	0.77
Alive & at home	60	51	58	0.70
n	195	49	244	
OVERALL				
Dead	21	31	24	0.10
Dead or dependent	57	67	60	0.16
Alive & at home	63	53	60	0.14
n	1257	460	1717	

* Calculated as the sum of the individual predictions of outcome (generated for each patient by the study models) divided by the number of patients.

Table 7.15 Observed outcomes six months after admission per hospital

Hospital	Proportion observed with each outcome at six months					
	Dead		Dead or dependent		Alive & at home	
	%	n/d	%	n/d	%	n/d
Hospital A	**48**	251 / 521	**89**	406 / 457	**40**	184 / 461
Hospital B	**37**	273 / 746	**90**	504 / 560	**45**	253 / 564
Hospital C	**38**	195 / 520	**87**	367 / 420	**43**	183 / 425
Hospital D	**27**	147 / 551	**66**	303 / 460	**58**	257 / 446
Hospital E	**37**	142 / 386	**84**	273 / 326	**49**	163 / 333
Difference p (χ^2)	<0.0005		<0.0005		<0.0005	

n/d number of outcomes / number of patients with follow up data

Table 7.16a Crude and adjusted case fatality at six months shown as W and Ws scores and as odds ratios

Model used to adjust for casemix	Patients	Mortality		Excess mortality per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds Ratio (95% CI)
Unadjusted								p < 0.0005
Hosp A	521	250	193	**11.0**	(6.8, 15.1)	**11.0**	(6.8, 15.1)	reference
Hosp B	746	273	276	**-0.4**	(-3.9, 3.1)	**-0.4**	(-3.9, 3.1)	**0.63** (0.50-0.79)
Hosp C	520	195	193	**0.5**	(-3.6, 4.6)	**0.5**	(-3.6, 4.6)	**0.65** (0.51-0.83)
Hosp D	551	147	204	**-10.3**	(-14.4, -6.3)	**-10.3**	(-14.4, -6.3)	**0.39** (0.31-0.51)
Hosp E	386	142	143	**-0.2**	(-5.0, 4.6)	**-0.2**	(-5.0, 4.6)	**0.63** (0.48-0.83)
Age, sex, social class								p < 0.0005
Hosp A	521	250	201	**9.4**	(5.4, 13.4)	**9.0**	(4.9, 13.1)	reference
Hosp B	746	273	267	**0.8**	(-2.5, 4.1)	**0.9**	(-3.2, 5.0)	**0.69** (0.51-0.93)
Hosp C	520	195	200	**-0.9**	(-4.9, 3.1)	**0.2**	(-4.0, 4.3)	**0.62** (0.47-0.82)
Hosp D	551	147	190	**-7.8**	(-11.6, -4.1)	**-8.1**	(-12.2, -3.9)	**0.41** (0.31-0.54)
Hosp E	386	142	149	**-1.8**	(-6.5, 2.8)	**-2.0**	(-6.1, 2.1)	**0.60** (0.46-0.80)
Study model								p = 0.047
Hosp A	521	250	225	**4.8**	(1.4, 8.3)	**4.6**	(2.1, 7.1)	reference
Hosp B	746	273	281	**-1.1**	(-3.9, 1.8)	**-1.2**	(-3.7, 1.4)	**0.70** (0.53-0.92)
Hosp C	520	195	206	**-2.0**	(-5.5, 1.5)	**-1.5**	(-4.1, 1.0)	**0.66** (0.49-0.89)
Hosp D	551	147	151	**-0.8**	(-3.7, 2.1)	**-1.7**	(-4.2, 0.9)	**0.70** (0.50-0.96)
Hosp E	386	142	144	**-0.6**	(-4.5, 3.3)	**-0.4**	(-2.9, 2.2)	**0.72** (0.51-1.00)

Table 7.16a cont.

Model used to adjust for casemix	Patients	Mortality		Excess mortality per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds Ratio (95% CI)
Urinary incontinence								p < 0.0005
Hosp A	521	250	197	**10.1**	(6.3, 13.9)	**10.0**	(3.7, 16.4)	reference
Hosp B	746	273	285	**-1.6**	(-4.7, 1.6)	**-1.5**	(-7.8, 4.9)	**0.55** (0.42-0.71)
Hosp C	520	195	198	**-0.5**	(-4.3, 3.3)	**-0.5**	(-6.9, 5.8)	**0.58** (0.44-0.76)
Hosp D	551	147	176	**-5.2**	(-8.7, -1.7)	**-5.5**	(-11.9, 0.8)	**0.44** (0.33-0.58)
Hosp E	386	142	152	**-2.5**	(-6.9, 2.0)	**-2.6**	(-9.0, 3.8)	**0.52** (0.39-0.71)
Study model + u.incontinence								p = 0.012
Hosp A	521	250	221	**5.5**	(2.1, 8.9)	**5.2**	(2.7, 7.7)	reference
Hosp B	746	273	282	**-1.2**	(-4.0, 1.6)	**-1.1**	(-3.6, 1.3)	**0.65** (0.49-0.86)
Hosp C	520	195	204	**-1.8**	(-5.2, 1.6)	**-3.5**	(-6.0, -1.0)	**0.62** (0.46-0.85)
Hosp D	551	147	152	**-1.0**	(-3.9, 1.9)	**-1.2**	(-3.7, 1.2)	**0.65** (0.46-0.90)
Hosp E	386	142	147	**-1.3**	(-5.2, 2.6)	**-1.3**	(-3.7, 1.2)	**0.64** (0.46-0.90)
Study model + social class								p = 0.074
Hosp A	521	250	228	**4.2**	(0.8, 7.7)	**4.0**	(1.5, 6.5)	reference
Hosp B	746	273	275	**-0.2**	(-3.1, 2.6)	**-0.2**	(-2.7, 2.4)	**0.72** (0.50-1.03)
Hosp C	520	195	206	**-2.1**	(-5.6, 1.4)	**-1.4**	(-3.9, 1.1)	**0.65** (0.47-0.90)
Hosp D	551	147	151	**-0.7**	(-3.7, 2.2)	**-1.6**	(-4.1, 0.9)	**0.71** (0.51-1.00)
Hosp E	386	142	147	**-1.4**	(-5.3, 2.6)	**-1.7**	(-4.3, 0.8)	**0.71** (0.51-0.98)

Table 7.16a cont.

Model used to adjust for casemix	Patients	Mortality		Excess mortality per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds Ratio (95% CI)
All casemix variables ‡								p = 0.013
Hosp A	478	227	204	**4.9**	(1.7, 8.1)	**5.0**	(2.6, 7.3)	reference
Hosp B	700	253	256	**-0.4**	(-3.1, 2.3)	**-0.8**	(-3.1, 1.5)	**0.60** (0.40-0.91)
Hosp C	494	182	192	**-2.1**	(-5.3, 1.1)	**-1.9**	(-4.1, 0.4)	**0.55** (0.38-0.80)
Hosp D	515	135	140	**-0.9**	(-3.8, -1.9)	**-0.8**	(-3.0, 1.5)	**0.60** (0.41-0.89)
Hosp E	347	124	129	**-1.5**	(-5.4, 2.4)	**-1.7**	(-4.1, 0.6)	**0.61** (0.42-0.89)

* The predicted number of deaths is calculated as the sum of the individual predicted probabilities of death generated for each patient by the prognostic model. For the unadjusted data, the latter is simply the observed proportion of deaths in the whole study cohort.

† Odds ratios and p values derived from multiple logistic regression analysis (see 3.9.3); Hospital A set as reference category

‡ This model took into account all the baseline variables listed in Table 7.5 except for stroke pathological sub-type. Missing values were not replaced except for the study model, urinary incontinence and social deprivation variables, hence n = 2534

Table 7.16b Crude and adjusted death or dependency at six months shown as W and Ws scores and as odds ratios

Model used to adjust for casemix	Patients	Number dead or dependent		Excess number dead or dependent per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds Ratio (95% CI)
Unadjusted								p < 0.0005
Hosp A	457	406	381	**5.5**	(2.1, 8.9)	**5.5**	(2.1, 8.9)	reference
Hosp B	560	504	467	**6.6**	(3.6, 9.7)	**6.6**	(3.6, 9.7)	**1.13** (0.76-1.69)
Hosp C	420	367	350	**4.0**	(0.5, 7.6)	**4.0**	(0.5, 7.6)	**0.87** (0.58-1.31)
Hosp D	460	303	384	**-17.5**	(-20.9, -14.3)	**-17.5**	(-20.9, -14.3)	**0.24** (0.17-0.34)
Hosp E	326	273	272	**0.4**	(-3.7, 4.4)	**0.4**	(-3.7, 4.4)	**0.65** (0.43-0.98)
Age, sex, social class								p < 0.0005
Hosp A	457	406	377	**6.3**	(3.0, 9.7)	**5.6**	(2.1, 9.2)	reference
Hosp B	560	504	488	**2.8**	(0.2, 5.5)	**4.2**	(0.8, 7.6)	**0.94** (0.58-1.54)
Hosp C	420	367	358	**2.3**	(-1.0, 5.6)	**2.5**	(-1.1, 6.0)	**0.75** (0.48-1.17)
Hosp D	460	303	361	**-12.7**	(-16.2, -9.2)	**-12.8**	(-16.9, -8.8)	**0.24** (0.16-0.35)
Hosp E	326	273	269	**1.2**	(-2.8, 5.2)	**1.5**	(-2.1, 5.1)	**0.62** (0.41-0.96)
Study model								p < 0.0005
Hosp A	457	406	357	**10.7**	(8.0, 13.3)	**14.2**	(12.2, 16.9)	reference
Hosp B	560	504	420	**14.9**	(12.3, 17.6)	**16.6**	(14.7, 18.6)	**1.33** (0.85-2.07)
Hosp C	420	367	325	**10.0**	(7.1, 12.9)	**13.7**	(11.6, 15.8)	**0.89** (0.56-1.40)
Hosp D	460	303	272	**6.8**	(3.7, 10.0)	**2.9**	(0.9, 5.0)	**0.38** (0.26-0.57)
Hosp E	326	273	239	**10.4**	(6.8, 14.0)	**10.9**	(8.9, 12.9)	**0.72** (0.45-1.14)

Table 7.16b cont.

Model used to adjust for casemix	Patients	Number dead or dependent		Excess number dead or dependent per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds Ratio (95% CI)
Urinary incontinence								p < 0.0005
Hosp A	457	406	383	**5.0**	(1.9, 8.1)	**5.1**	(0.6, 9.7)	reference
Hosp B	560	504	473	**5.6**	(2.8, 8.4)	**6.1**	(1.5, 10.6)	**1.09** (0.72-1.66)
Hosp C	420	367	354	**3.1**	(-0.1, 6.4)	**3.3**	(-1.3, 7.8)	**0.83** (0.54-1.28)
Hosp D	460	303	368	**-14.0**	(-17.4, -10.6)	**-12.4**	(-16.9, -7.8)	**0.26** (0.18-0.38)
Hosp E	326	273	276	**-0.8**	(-4.5, 2.8)	**-1.1**	(-5.6, 3.4)	**0.57** (0.37-0.89)
Study model + u.incontinence								p < 0.0005
Hosp A	457	406	393	**2.8**	(0.0, 5.5)	**3.7**	(1.2, 6.2)	reference
Hosp B	560	504	476	**5.1**	(2.5, 7.6)	**5.8**	(3.3, 8.4)	**1.28** (0.82-2.00)
Hosp C	420	367	361	**1.6**	(-1.3, 4.5)	**2.7**	(0.2, 5.3)	**0.86** (0.54-1.36)
Hosp D	460	303	348	**-9.8**	(-13.2, -6.4)	**-8.2**	(-10.8, -5.5)	**0.36** (0.24-0.54)
Hosp E	326	273	275	**-0.7**	(-4.2, 2.7)	**-0.6**	(-3.1, 2.0)	**0.67** (0.42-1.06)
Study model + social class								p < 0.0005
Hosp A	457	406	391	**3.2**	(0.4, 6.0)	**3.9**	(1.4, 6.4)	reference
Hosp B	560	504	489	**2.7**	(0.3, 5.2)	**4.0**	(1.5, 6.4)	**0.97** (0.63-1.51)
Hosp C	420	367	364	**0.8**	(-2.0, 3.7)	**1.9**	(-0.7, 4.4)	**0.77** (0.49-1.21)
Hosp D	460	303	338	**-7.6**	(-11.1, -4.1)	**-6.5**	(-9.1, -3.9)	**0.41** (0.28-0.61)
Hosp E	326	273	272	**0.4**	(-3.2, 4.1)	**1.3**	(-1.2, 3.9)	**0.72** (0.45-1.14)

Table 7.16b cont.

Model used to adjust for casemix	Patients	Number dead or dependent		Excess number dead or dependent per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds Ratio (95% CI)
All casemix variables ‡								p < 0.0005
Hosp A	421	374	359	**3.7**	(0.8, 6.5)	**4.0**	(1.6, 6.3)	reference
Hosp B	527	474	460	**2.6**	(0.1, 5.1)	**4.0**	(1.7, 6.3)	**0.89** (0.55-1.44)
Hosp C	399	348	344	**1.0**	(-0.3, 2.2)	**1.5**	(-0.8, 3.9)	**0.72** (0.44-1.17)
Hosp D	434	284	318	**-7.7**	(-11.2, -4.3)	**-6.3**	(-8.8, -3.8)	**0.35** (0.22-0.53)
Hosp E	292	244	243	**0.2**	(-3.4, 3.8)	**-0.3**	(-2.8, 2.1)	**0.64** (0.38-1.06)

* The predicted number of patients dead or dependent is calculated as the sum of the individual predicted probabilities of death or dependency generated for each patient by the prognostic model. For the unadjusted data, the latter is simply the observed proportion of patients dead or dependent in the whole study cohort.

† Odds ratios and p values derived from multiple logistic regression analysis (see 3.9.3); Hospital A set as reference category

‡ This model took into account all the baseline variables listed in Table 7.5 except for stroke pathological sub-type. Missing values were not replaced except for the study model, urinary incontinence and social deprivation variables, hence n = 2073

Table 7.16c Crude and adjusted 'alive & at home' at six months shown as W and Ws scores and as odds ratios

Model used to adjust for casemix	Patients	Number alive & at home		Excess number alive & at home per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds ratio (95% CI)
Unadjusted								p < 0.0005
Hosp A	461	184	215	**-6.7**	(-11.3, -2.2)	**-6.7**	(-11.3, -2.2)	reference
Hosp B	564	253	263	**-1.8**	(-5.9, 2.3)	**-1.8**	(-5.9, 2.3)	**1.22** (0.95-1.57)
Hosp C	425	183	198	**-3.6**	(-8.3, 1.1)	**-3.6**	(-8.3, 1.1)	**1.14** (0.87-1.49)
Hosp D	446	257	208	**11.0**	(6.3, 15.6)	**11.0**	(6.3, 15.6)	**2.05** (1.57-2.67)
Hosp E	333	163	155	**2.3**	(-3.1, 7.6)	**2.3**	(-3.1, 7.6)	**1.44** (1.09-1.92)
Age, sex, social class								p = 0.0001
Hosp A	461	184	214	**-6.4**	(-10.7, -2.2)	**-6.4**	(-10.1, -2.8)	reference
Hosp B	564	253	262	**-1.6**	(-5.5, 2.2)	**-1.6**	(-5.3, 2.0)	**1.18** (0.84-1.65)
Hosp C	425	183	185	**-0.4**	(-4.8, 4.0)	**-0.4**	(-4.1, 3.2)	**1.31** (0.97-1.78)
Hosp D	446	257	225	**7.1**	(2.9, 11.4)	**7.8**	(4.1, 11.5)	**2.01** (1.49-2.72)
Hosp E	333	163	154	**2.6**	(-2.4, 7.6)	**2.5**	(0.7, 4.4)	**1.52** (1.12-2.07)
Study model								p = 0.574
Hosp A	461	184	186	**-0.4**	(-3.8, 3.1)	**-0.5**	(-2.8, 1.8)	reference
Hosp B	564	253	247	**1.0**	(-2.2, 4.2)	**1.0**	(-1.3, 3.3)	**1.10** (0.81-1.51)
Hosp C	425	183	175	**1.9**	(-1.9, 5.6)	**1.5**	(-0.8, 3.8)	**1.15** (0.83-1.61)
Hosp D	446	257	259	**-0.6**	(-3.9, 2.8)	**0.0**	(-2.3, 2.3)	**1.02** (0.75-1.45)
Hosp E	333	163	150	**3.8**	(-0.3, 8.0)	**3.3**	(1.0, 5.6)	**1.33** (0.93-1.90)

Table 7.16c cont.

Model used to adjust for casemix	Patients	Number alive & at home		Excess number alive & at home per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds ratio (95% CI)
Urinary incontinence								p = 0.0006
Hosp A	461	184	212	**-6.1**	(-10.0, -2.3)	**-6.1**	(-12.1, -0.1)	reference
Hosp B	564	253	252	**0.2**	(-3.3, 3.7)	**0.2**	(-5.8, 6.2)	**1.44** (1.07-1.93)
Hosp C	425	183	191	**-1.9**	(-6.0, 2.1)	**-1.9**	(-7.9, 4.1)	**1.27** (0.93-1.75)
Hosp D	446	257	237	**4.5**	(0.7, 8.4)	**4.9**	(-1.1, 10.9)	**1.84** (1.35-2.52)
Hosp E	333	163	148	**4.5**	(0.0, 9.1)	**4.5**	(-1.5, 10.5)	**1.82** (1.30-2.55)
Study model + u.incontinence								p = 0.249
Hosp A	461	184	195	**-2.4**	(-5.8, 1.0)	**-2.2**	(-4.4, 0.1)	reference
Hosp B	564	253	250	**0.6**	(-2.5, 3.7)	**0.9**	(-1.4, 3.2)	**1.24** (0.89-1.73)
Hosp C	425	183	182	**0.4**	(-3.3, 4.0)	**0.5**	(-1.8, 2.8)	**1.22** (0.86-1.73)
Hosp D	446	257	262	**-1.2**	(-4.5, 2.1)	**-0.8**	(-3.1, 1.5)	**1.08** (0.75-1.55)
Hosp E	333	163	151	**3.5**	(-0.6, 7.6)	**3.2**	(0.9, 5.6)	**1.52** (1.04-2.21)
Study model + social class								p = 0.403
Hosp A	461	184	192	**-1.7**	(-5.3, 1.8)	**-1.6**	(-3.9, 0.8)	reference
Hosp B	564	253	256	**-0.6**	(-3.9, 2.7)	**-0.6**	(-3.0, 1.7)	**1.03** (0.69-1.54)
Hosp C	425	183	176	**1.6**	(-2.2, 5.4)	**1.7**	(-0.7, 4.0)	**1.26** (0.88-1.80)
Hosp D	446	257	262	**-1.0**	(-4.5, 2.4)	**-1.2**	(-3.6, 1.1)	**1.02** (0.71-1.47)
Hosp E	333	163	154	**2.8**	(-1.5, 7.0)	**2.0**	(-0.3, 4.4)	**1.34** (0.94-1.92)

Table 7.16c cont.

Model used to adjust for casemix	Patients	Number alive & at home		Excess number alive & at home per 100 admissions after adjusting for casemix				Comparison with Hospital A †
		Observed	Predicted * (for w score)	W score	95%CI	Ws Score	95%CI	Odds ratio (95% CI)
All casemix variables ‡								p = 0.334
Hosp A	425	173	182	**-2.2**	(-5.5, 1.1)	**-1.4**	(-3.3, 0.5)	reference
Hosp B	527	243	244	**-0.1**	(-3.2, 2.9)	**-0.3**	(-2.4, 1.8)	**1.20** (0.76-1.91)
Hosp C	404	179	173	**1.4**	(-2.1, 4.9)	**1.4**	(-0.7, 3.5)	**1.38** (0.91-2.09)
Hosp D	421	249	253	**-0.9**	(-4.2, 2.3)	**-1.0**	(-3.1, 1.1)	**1.10** (0.72-1.68)
Hosp E	299	150	142	**2.7**	(-1.4, 6.9)	**2.5**	(0.3, 4.6)	**1.49** (0.98-2.26)

* The predicted number of patients alive & at home is calculated as the sum of the individual predicted probabilities of being alive & at home generated for each patient by the prognostic model. For the unadjusted data, the latter is simply the observed proportion of patients alive & at home in the whole study cohort.

† The p value relates to the significance of a hospital term (Hospital A set as reference category) forced in to each regression analysis

‡ This model took into account all the baseline variables listed in Table 7.5 except for stroke pathological sub-type. Missing values were not replaced except for the study model, urinary incontinence and social deprivation variables, hence n = 2076

Table 7.17 Comparison of adjusted outcomes: primary SOP method of measuring outcome versus various alternatives

Outcome at six months	**Adjusted odds ratios** (95% CI)				
	Primary method of measuring outcome	Replacement of missing predictor values with the most optimistic option	Recoding of cases 'alive by linked survival data but dead by SMR1 or GP' as dead	Exclusion of cases in which there was difficulty in extracting data	Exclusion of cases in which stroke onset occurred whilst admitted for another disorder
Death					
Hosp. B v A	0.70 (0.53 - 0.92)	0.70 (0.53 - 0.93)	0.67 (0.50 - 0.88)	0.70 (0.52 - 0.94)	0.65 (0.49 - 0.87)
Hosp. C v A	0.66 (0.49 - 0.89)	0.65 (0.48 - 0.88)	0.62 (0.46 - 0.84)	0.65 (0.47 - 0.89)	0.64 (0.47 - 0.87)
Hosp. D v A	0.70 (0.51 - 0.96)	0.72 (0.52 - 0.99)	0.68 (0.49 - 0.93)	0.72 (0.51 - 1.02)	0.63 (0.45 - 0.88)
Hosp. E v A	0.72 (0.51 - 1.00)	0.72 (0.52 - 1.00)	0.68 (0.48 - 0.95)	0.72 (0.48 - 0.98)	0.71 (0.51 - 1.00)
p	0.047	0.047	0.016	**0.075**	0.015
n	2724	2724	2724	2497	2542
Death or dependency					
Hosp. B v A	1.33 (0.85 - 2.07)	1.33 (0.86 - 2.07)	1.31 (0.84 - 2.04)	1.34 (0.85 - 2.12)	1.33 (0.85 - 2.08)
Hosp. C v A	0.89 (0.56 - 1.40)	0.88 (0.56 - 1.39)	0.88 (0.56 - 1.39)	0.93 (0.58 - 1.49)	0.86 (0.54 - 1.36)
Hosp. D v A	0.38 (0.26 - 0.57)	0.40 (0.27 - 0.59)	0.38 (0.26 - 0.57)	0.39 (0.26 - 0.60)	0.38 (0.25 - 0.56)
Hosp. E v A	0.72 (0.45 - 1.14)	0.72 (0.46 - 1.15)	0.71 (0.45 - 1.13)	0.70 (0.43 - 1.12)	0.70 (0.44 - 1.12)
p	< 0.00005	< 0.00005	< 0.00005	< 0.00005	< 0.00005
n	2223	2223	2242	2042	2052
Alive & at home					
Hosp. B v A	1.10 (0.81 - 1.51)	1.09 (0.80 - 1.50)	1.13 (0.83 - 1.55)	1.06 (0.75 - 1.48)	1.15 (0.83 - 1.59)
Hosp. C v A	1.15 (0.83 - 1.61)	1.17 (0.84 - 1.63)	1.18 (0.85 - 1.65)	1.11 (0.78 - 1.58)	1.20 (0.86 - 1.69)
Hosp. D v A	1.02 (0.73 - 1.45)	0.98 (0.69 - 1.38)	1.04 (0.74 - 1.47)	0.94 (0.65 - 1.37)	1.08 (0.76 - 1.54)
Hosp. E v A	1.33 (0.93 - 1.90)	1.32 (0.92 - 1.89)	1.36 (0.95 - 1.94)	1.27 (0.87 - 1.85)	1.38 (0.96 - 1.98)
p	0.574	0.483	0.495	0.634	0.503
n	2229	2229	2248	2050	2059

Table 7.18 The impact of adding stroke pathological sub-type (haemorrhage vs. not) to the model that adjusts for all measured casemix

Outcome at six months	Odds ratio of each outcome (95% CI)	
	Adjusted for all casemix data EXCLUDING stroke pathological subtype	Adjusted for all casemix data INCLUDING stroke pathological subtype
Death		
Hosp B vs A	0.60 (0.40-0.91)	0.61 (0.41-0.93)
Hosp C vs A	0.55 (0.40-0.91)	0.58 (0.40-0.84)
Hosp D vs A	0.60 (0.41-0.89)	0.65 (0.44-0.95)
Hosp E vs A	0.61 (0.42-0.89)	0.64 (0.44-0.94)
p value (hospital term)	0.013	0.031
Haemorrhagic stroke term		0.79 (0.56-1.12)
p value		0.183
Death or dependency		
Hosp B vs A	0.87 (0.50-1.53)	1.02 (0.61-1.72)
Hosp C vs A	0.69 (0.41-1.16)	0.78 (0.47-1.29)
Hosp D vs A	0.31 (0.19-0.48)	0.34 (0.22-0.52)
Hosp E vs A	0.65 (0.39-1.09)	0.64 (0.38-1.07)
p value (hospital term)	<0.00005	<0.00005
Haemorrhagic stroke term		0.62 (0.39-0.97)
p value		0.037
Alive & at home		
Hosp B vs A	1.20 (0.76-1.91)	1.18 (0.74-1.87)
Hosp C vs A	1.38 (0.91-2.09)	1.35 (0.89-2.05)
Hosp D vs A	1.10 (0.72-1.68)	1.08 (0.71-1.65)
Hosp E vs A	1.49 (0.98-2.26)	1.45 (0.96-2.22)
p value (hospital term)	0.334	0.393
Haemorrhagic stroke term		1.40 (0.95-2.08)
p value		0.093

Table 7.19 Different methods of identifying patients admitted to hospital with stroke using routine data: impact on estimates of unadjusted case fatality and on their comparison between hospitals

Analysis based only on the patients admitted to hospital in the first six months of the study; Hospital A set as reference hospital

Hospital	Method used to identify patients with acute stroke from SMR1 data														
	Any CVD code listed as primary diagnosis			ISD method			DoH method			Modified ISD method			Review of medical record GOLD STANDARD		
	%dead	RR	n/d	%dead	RR	n/d	%dead	RR	n/d	%dead	RR	n/d	%dead	RR	n/d
Hosp A	**40**	Ref.	70/177	**43**	Ref.	66/153	**47**	Ref.	60/129	**44**	Ref.	65/147	**50**	Ref.	59/119
Hosp B	**30**	**0.75**	120/402	**38**	**0.88***	89/234	**39**	**0.83***	83/214	**38**	**0.86***	87/229	**37**	**0.74**	77/209
Hosp C	**37**	**0.92**	85/233	**43**	**0.99**	68/159	**44**	**0.95**	63/143	**43**	**0.98**	64/148	**47**	**0.96**	65/137
Hosp D	**21**	**0.53**	59/283	**26**	**0.60**	43/167	**25**	**0.54**	30/120	**25**	**0.57**	32/126	**26**	**0.53**	30/115
Hosp E	**31**	**0.77**	49/160	**36**	**0.83**	45/126	**36**	**0.77**	36/101	**36**	**0.81**	43/120	**40**	**0.81**	35/87

n/d Deaths at six months / all cases identified

* using these methods, the comparison of case fatality between Hospitals A & B appeared to be non-significant (ISD method RR 0.88; 95%CI: 0.69 to 1.13; DoH method RR 0.83; 95%CI: 0.65 to 1.07; Modified ISD method RR 0.86; 95%CI: 0.67 to 1.10) when in fact by the gold standard the difference in case fatality was just significant (RR 0.74; 0.58 to 0.96).

Figure 7.1 The identification of patients with acute non-SAH stroke

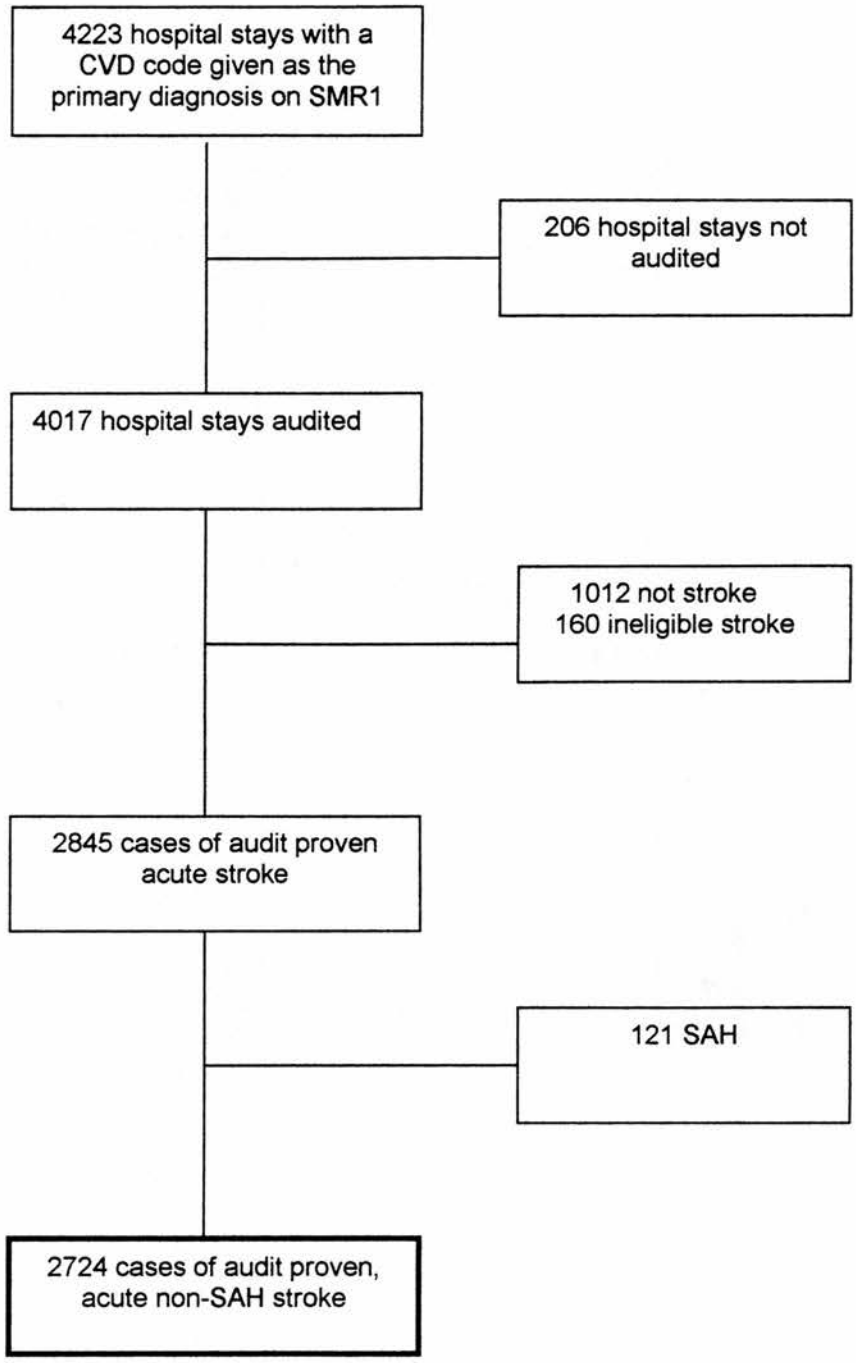
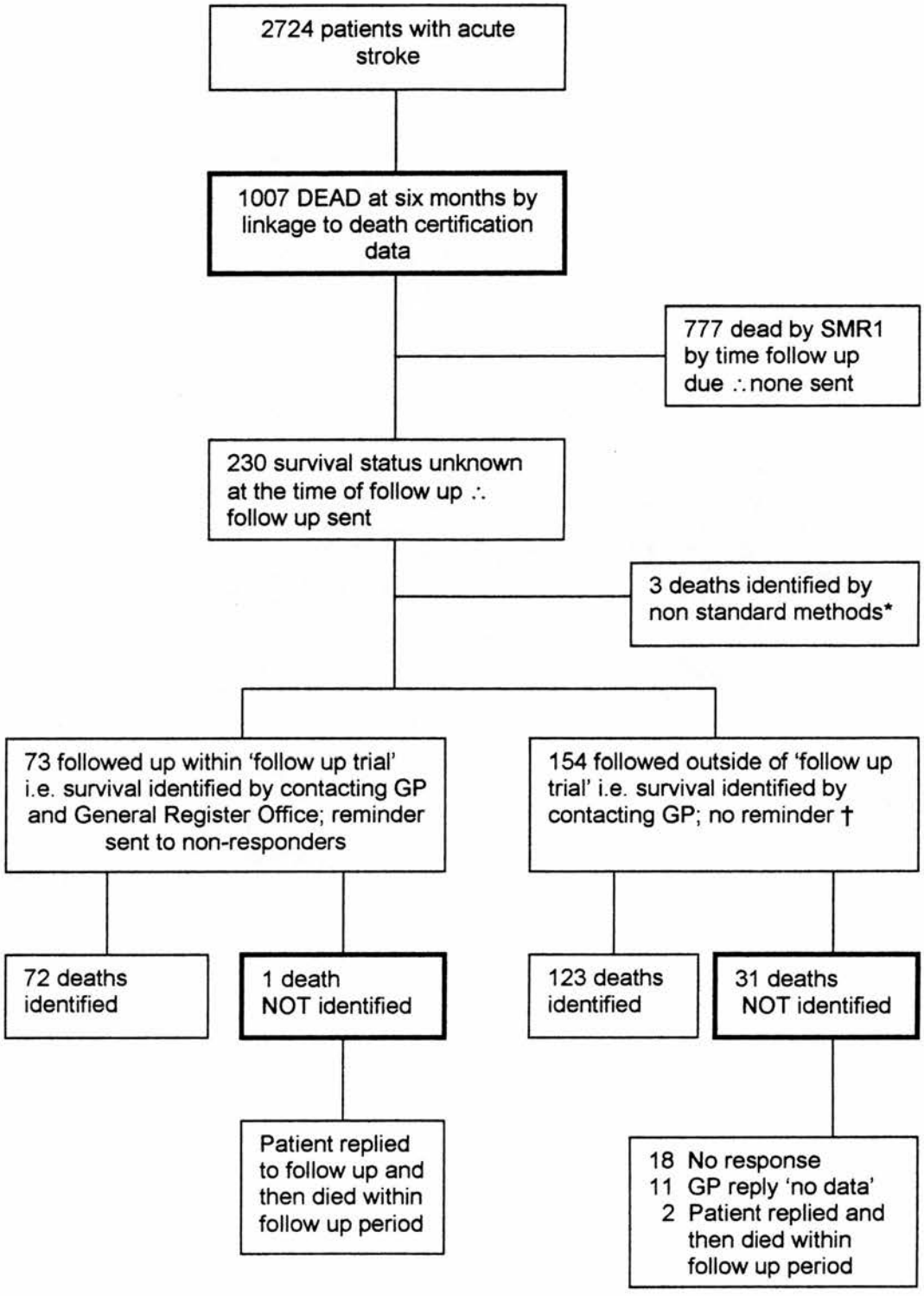


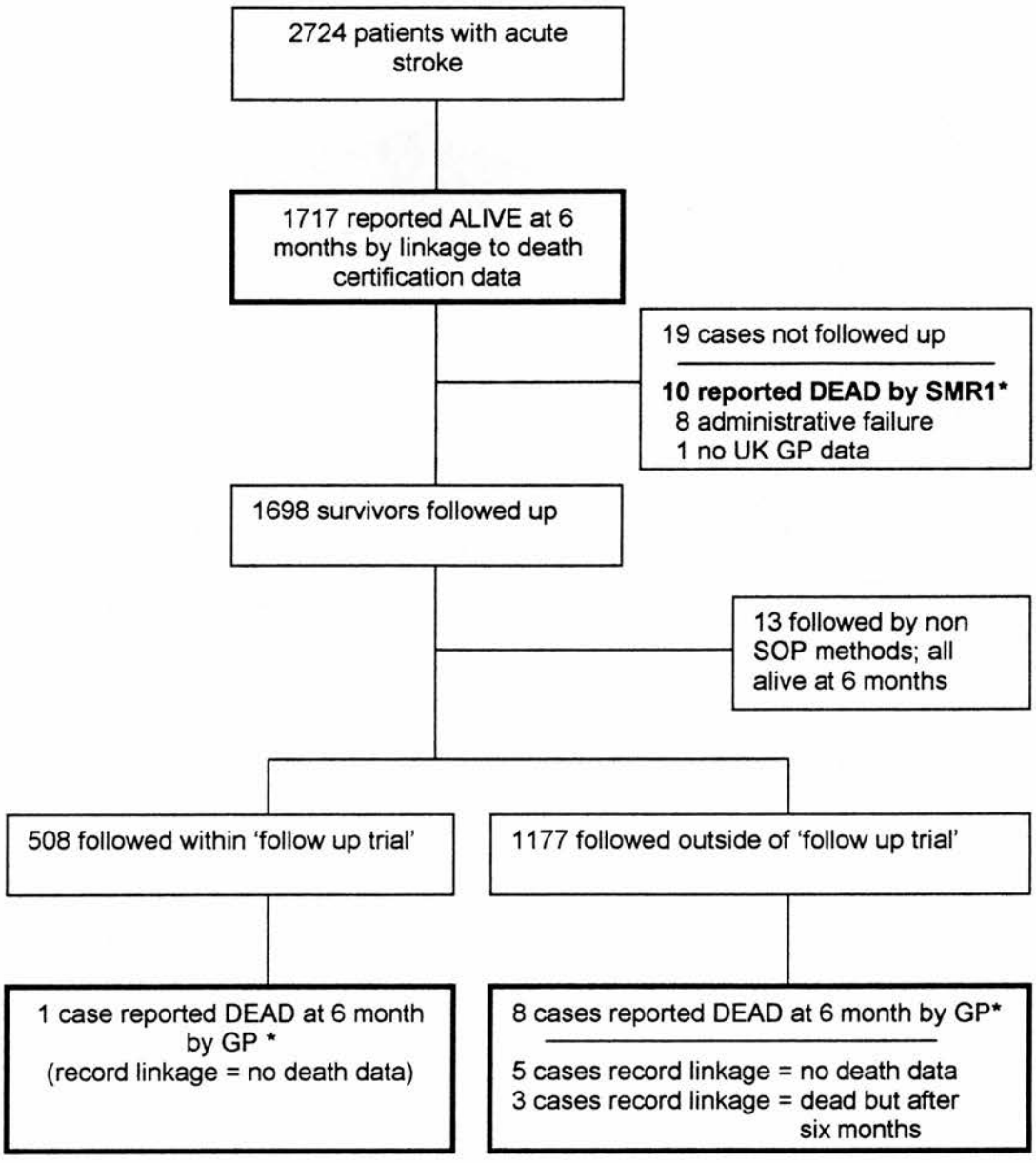
Figure 7.2 The success of the SOP 'hot-pursuit' method of follow up in identifying deaths at six months after admission (death certificate data used as gold standard)



* One via FOOD trial follow up, one via the Health Board & one via stroke register at Hospital D

† After the follow up trial a reminder was sent to non-responding GPs/patients but none of the GPs that reported death during this period (n = 4) required reminding

Figure 7.3 The success of the SOP 'hot-pursuit' method of follow up in identifying patients alive at six months after admission (death certificate data as gold standard)



* 10 + 1+ 8 = 19 cases (1%) of 1717 patients alive at six months by linkage to death certificate data were dead at six months according to SMR1/GP

* Distribution: 7 cases at Hosp A, 5 cases at Hosp B, 2 cases at Hosp C, 3 cases at Hosp D and 2 cases at Hosp E

Figure 7.4 The success in collecting follow up data from patients who were alive at six months after admission (according to death certificate data)

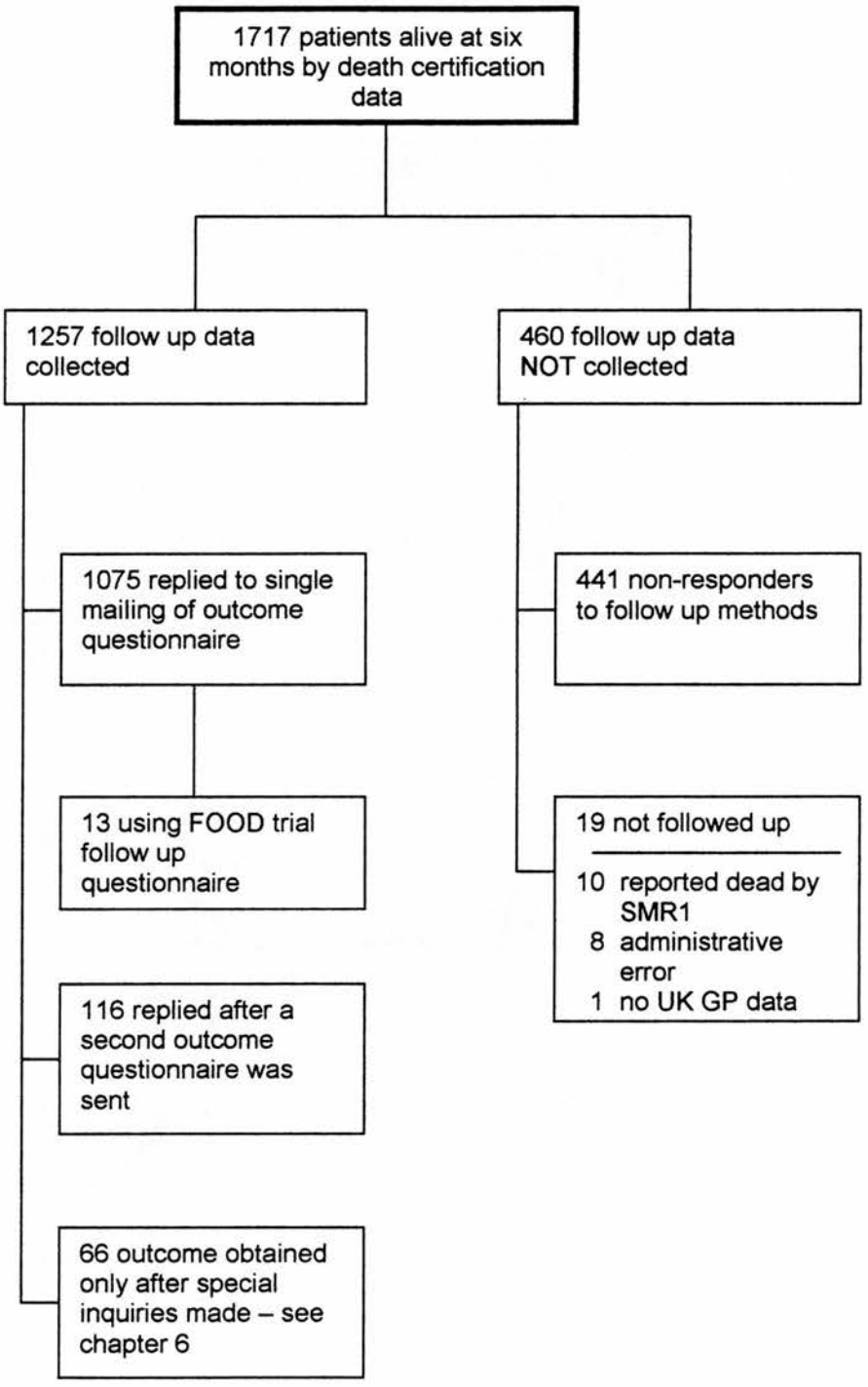
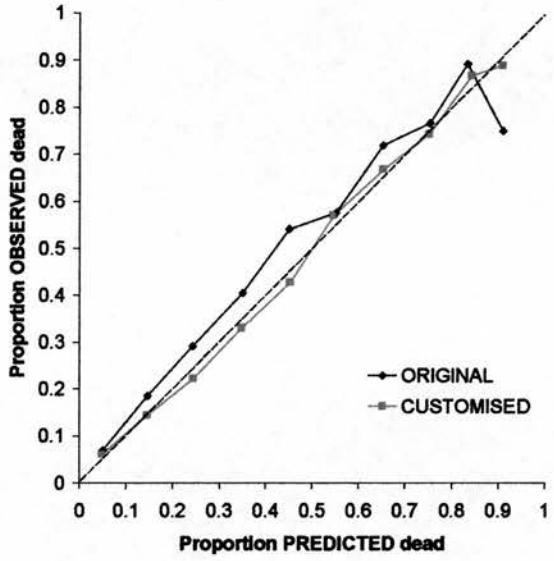


Figure 7.5 Calibration of the prognostic models in the SOP data set

1. Case fatality at six months



2. Dead or dependent at six months

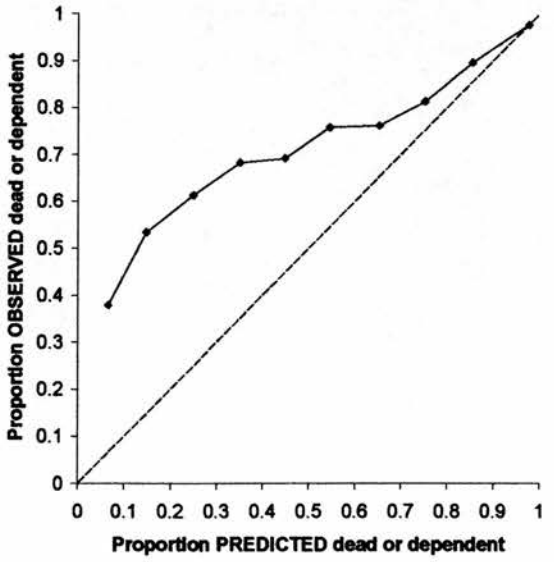
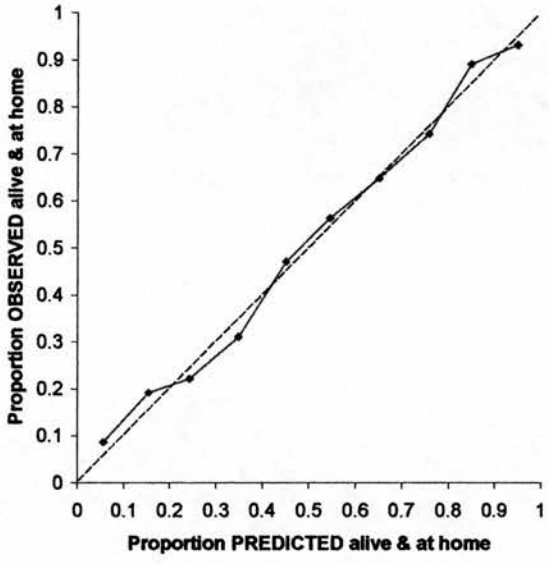


Figure 7.5 (cont.)

3. Alive and at home at six months



The dashed line represents the line of perfect calibration

Figure 7.6a W scores for case fatality at six months: unadjusted; adjusted for age, sex and social deprivation; and adjusted using the study model.

W score at Hospital A set to zero to facilitate comparisons between hospitals

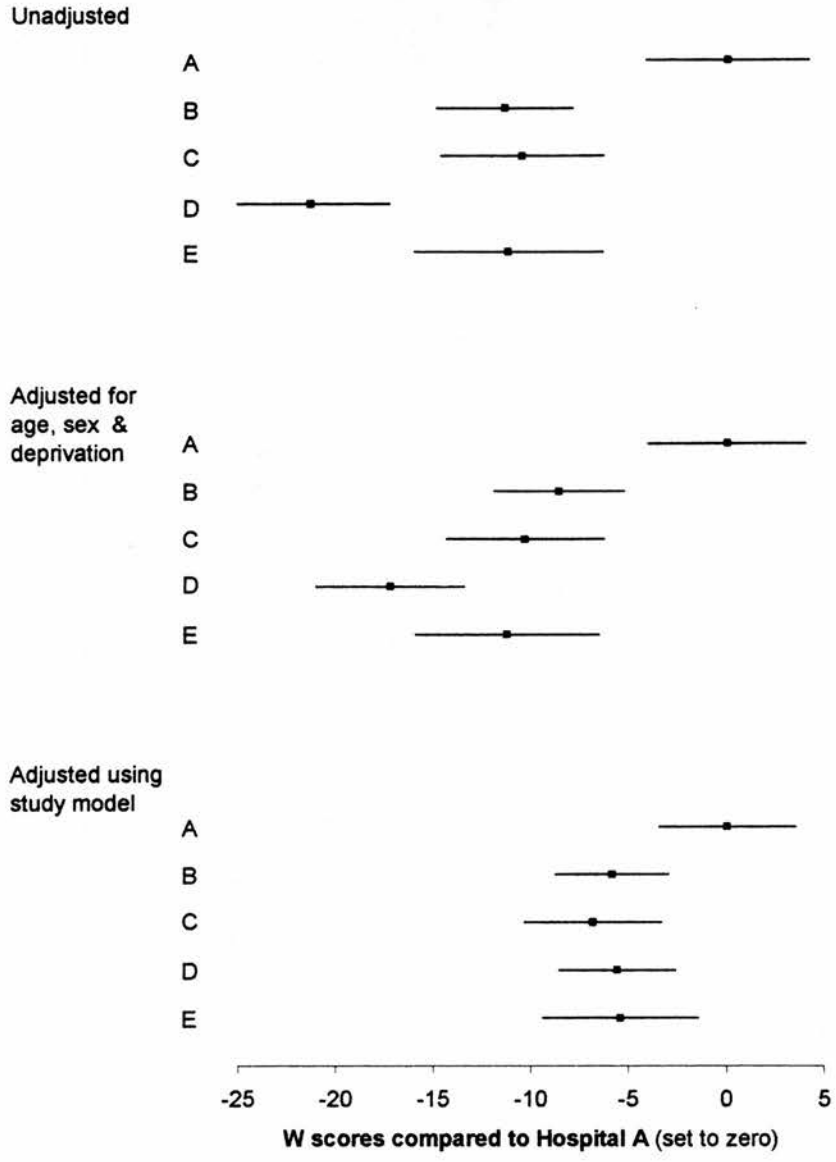


Figure 7.6b Ws scores for death or dependency at six months: unadjusted; adjusted for age, sex and social deprivation; and adjusted using the study model.

W score at Hospital A set to zero to facilitate comparisons between hospitals

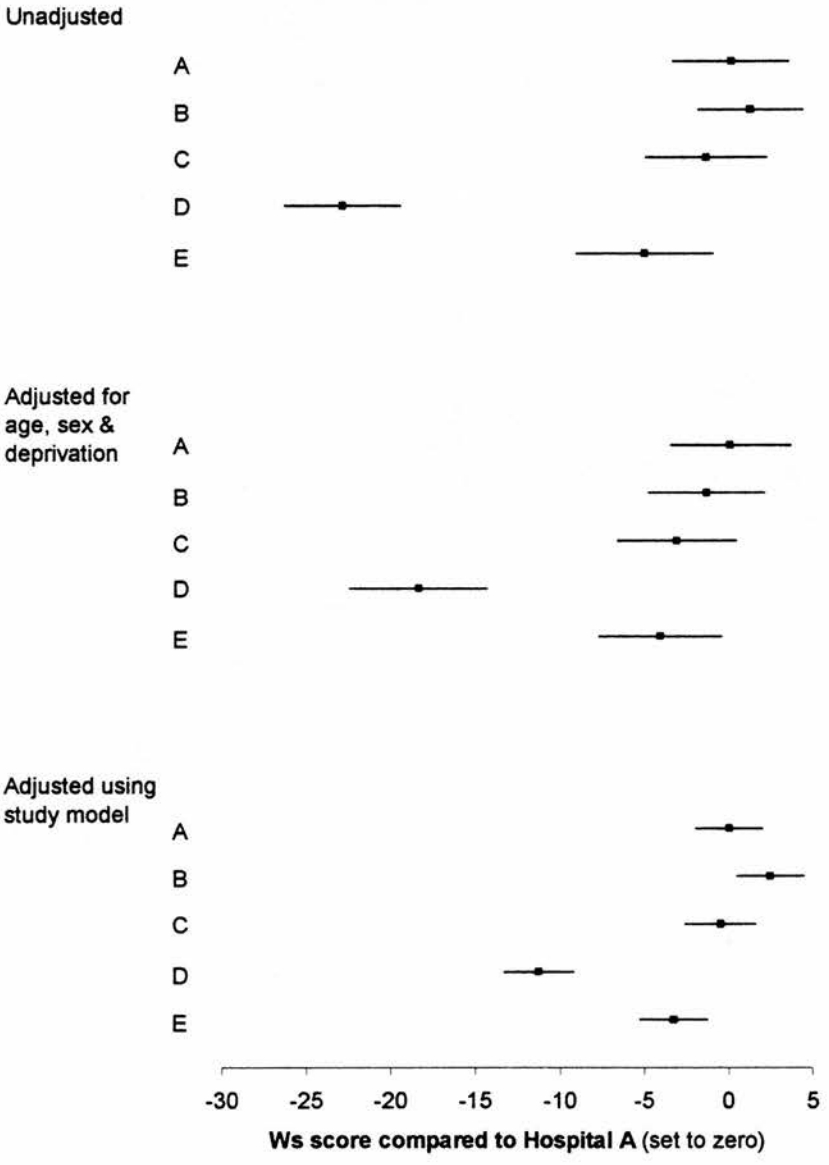


Figure 7.6c W scores for alive & at home at six months: unadjusted; adjusted for age, sex and social deprivation; and adjusted using the study model.

W score at Hospital A set to zero to facilitate comparisons between hospitals

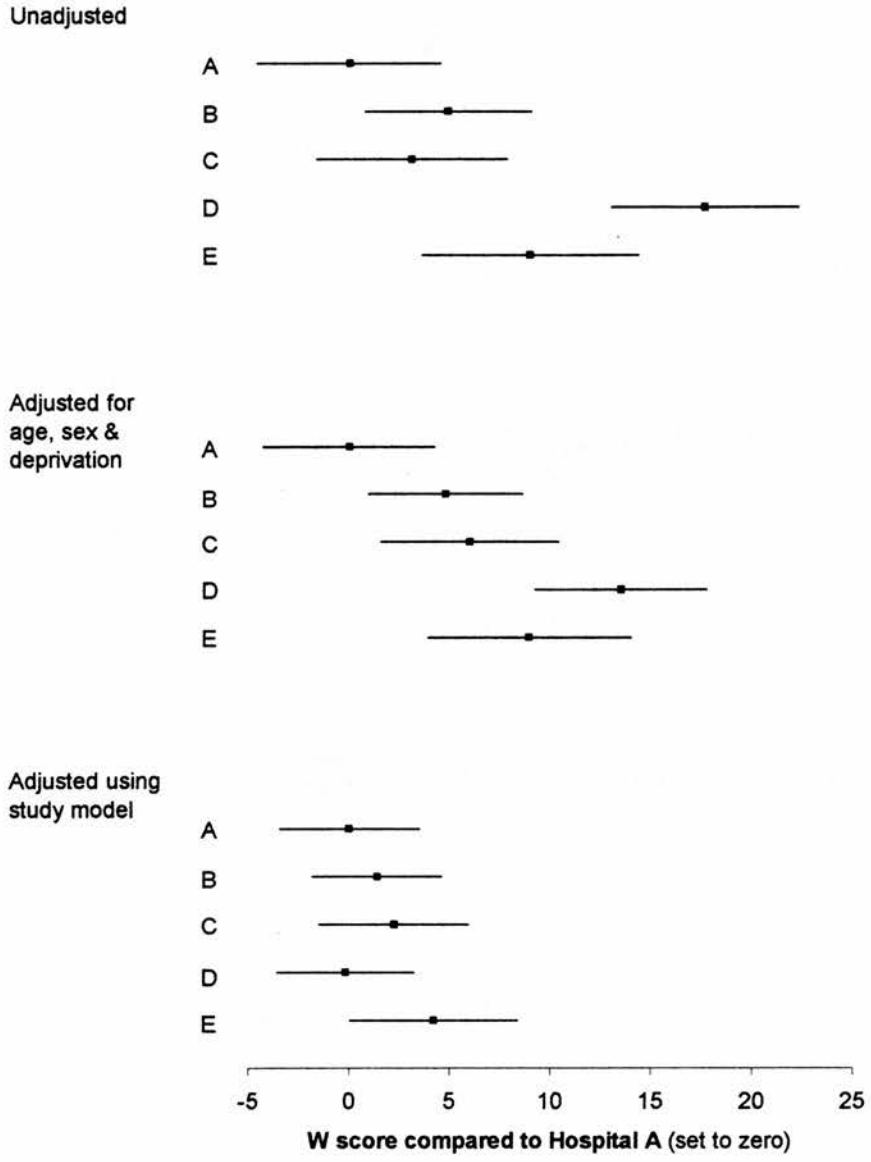


Figure 7.7a Alternative methods of adjusting case fatality at six months.

W score at Hospital A set to zero to facilitate comparisons between hospitals

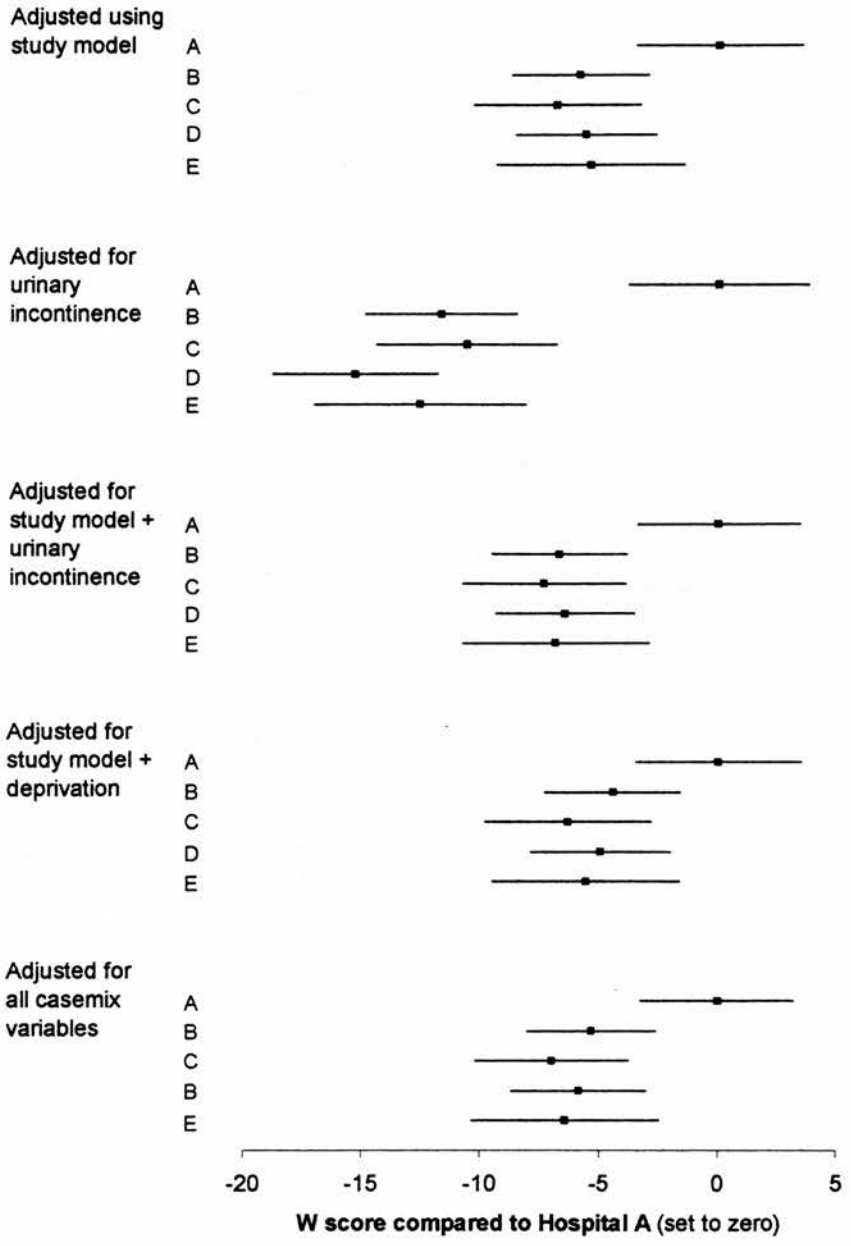


Figure 7.7b Alternative methods of adjusting death or dependency at six months.

W score at Hospital A set to zero to facilitate comparisons between hospitals

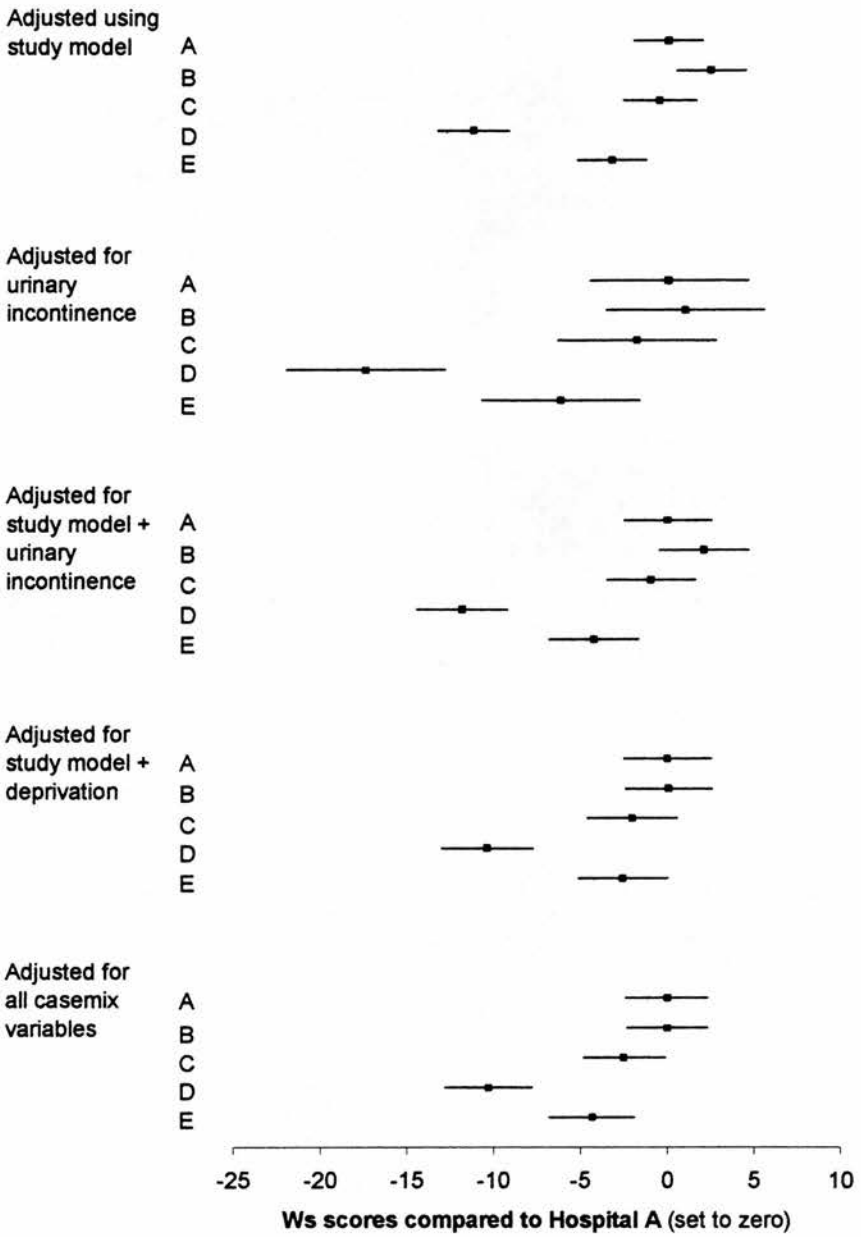


Figure 7.7c Alternative methods of adjusting alive & at home at six months.

W score at Hospital A set to zero to facilitate comparisons between hospitals

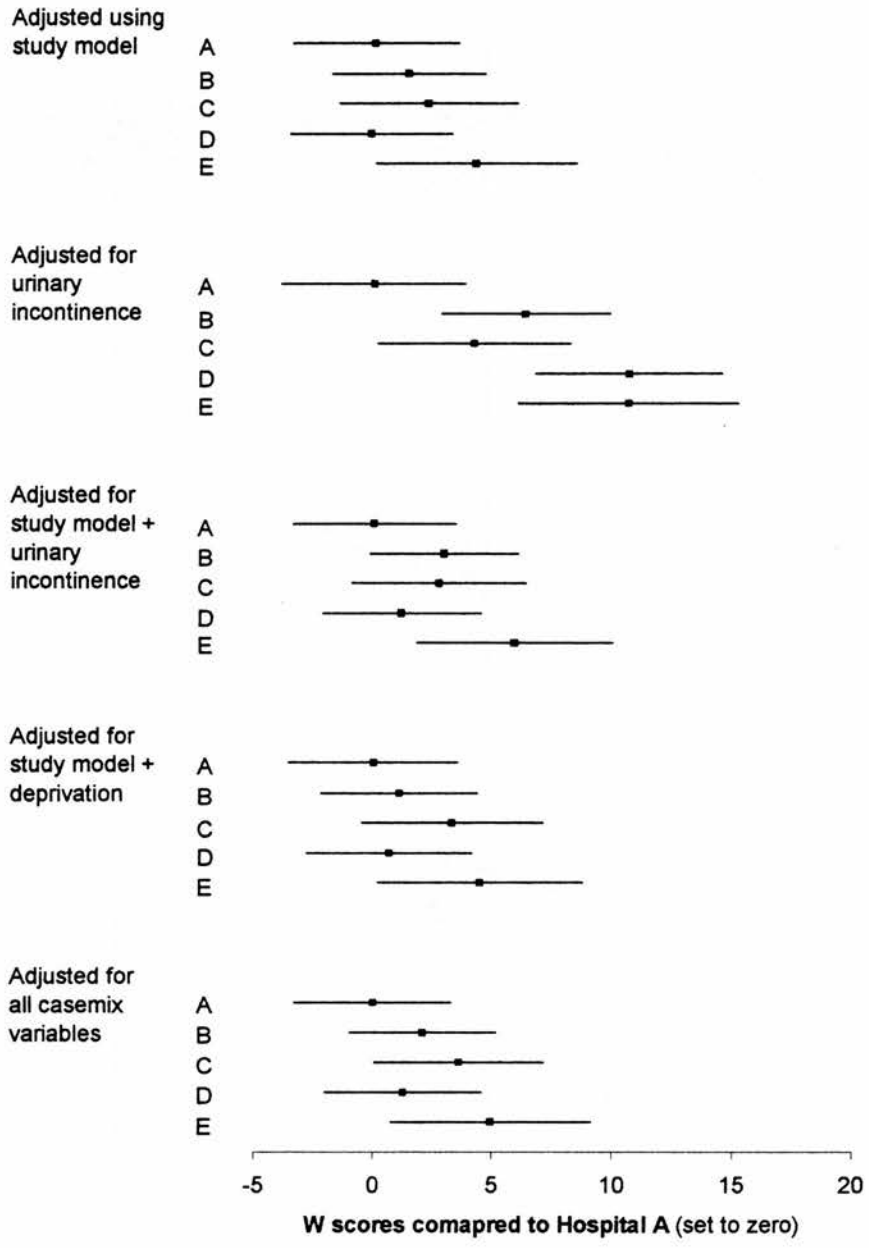


Figure 7.8 The logistic coefficients of the categories of social deprivation for death or dependency at six months plotted against each category of social deprivation

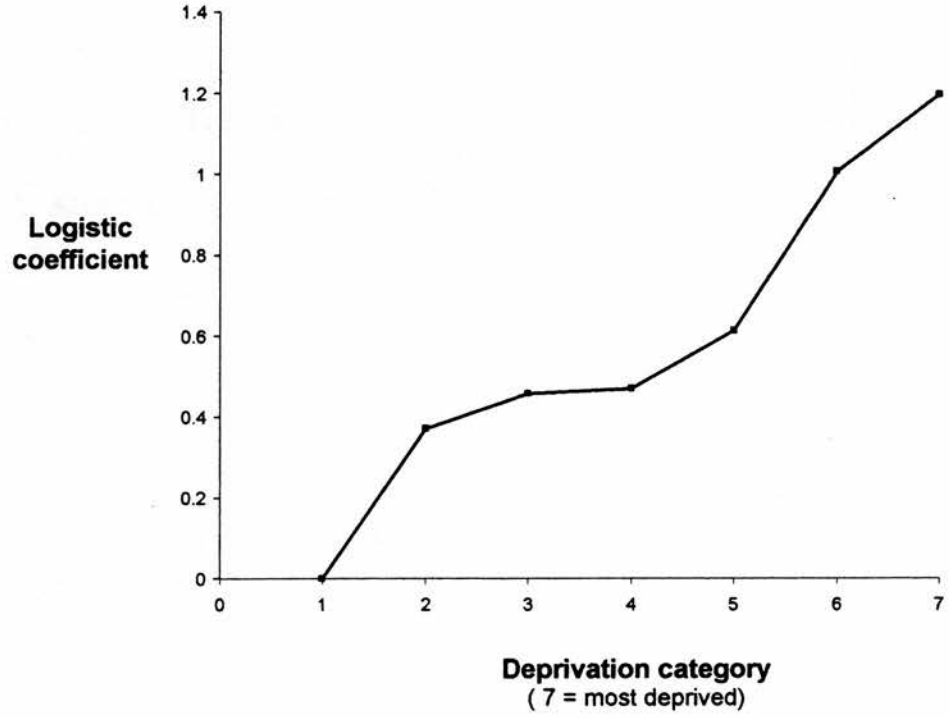


Figure 7.9a Comparison of W scores between hospitals for case fatality, adjusted using the study model, with the cohort divided into tertiles of predicted risk.

W score at Hospital A set to zero to facilitate comparisons between hospitals

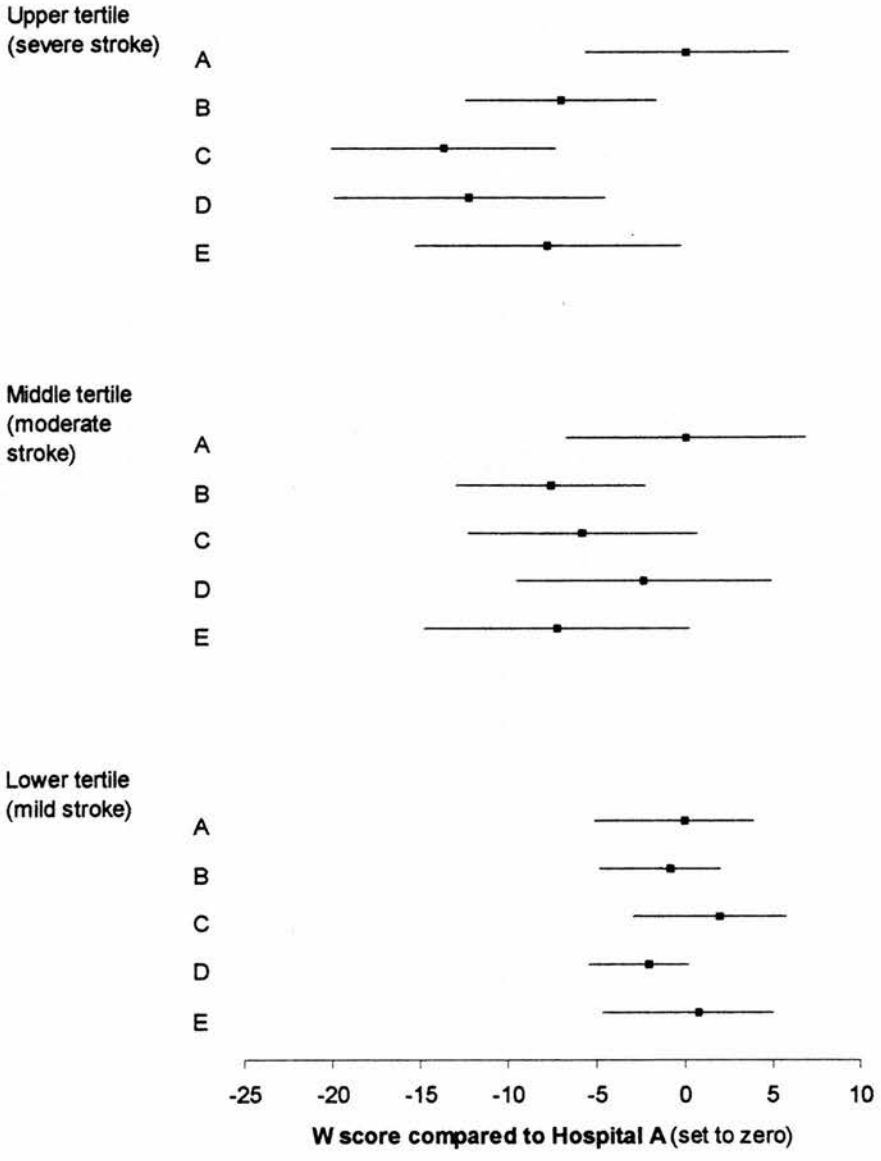


Figure 7.9b Comparison of Ws scores between hospitals for death or dependency, adjusted using the study model, with the cohort divided into tertiles of predicted risk.

W score at Hospital A set to zero to facilitate comparisons between hospitals

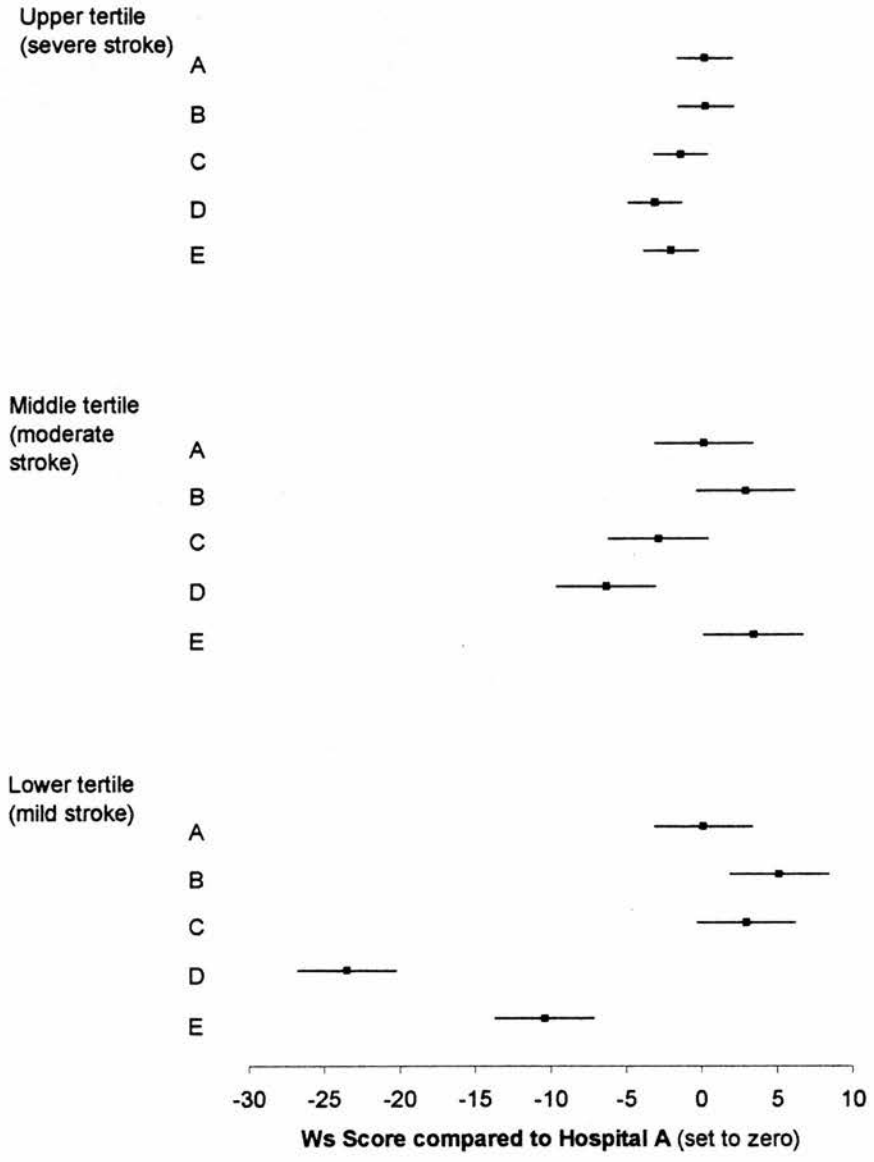


Figure 7.9c Comparison of W scores between hospitals for alive & at home, adjusted using the study model, with the cohort divided into tertiles of predicted risk.

W score at Hospital A set to zero to facilitate comparisons between hospitals

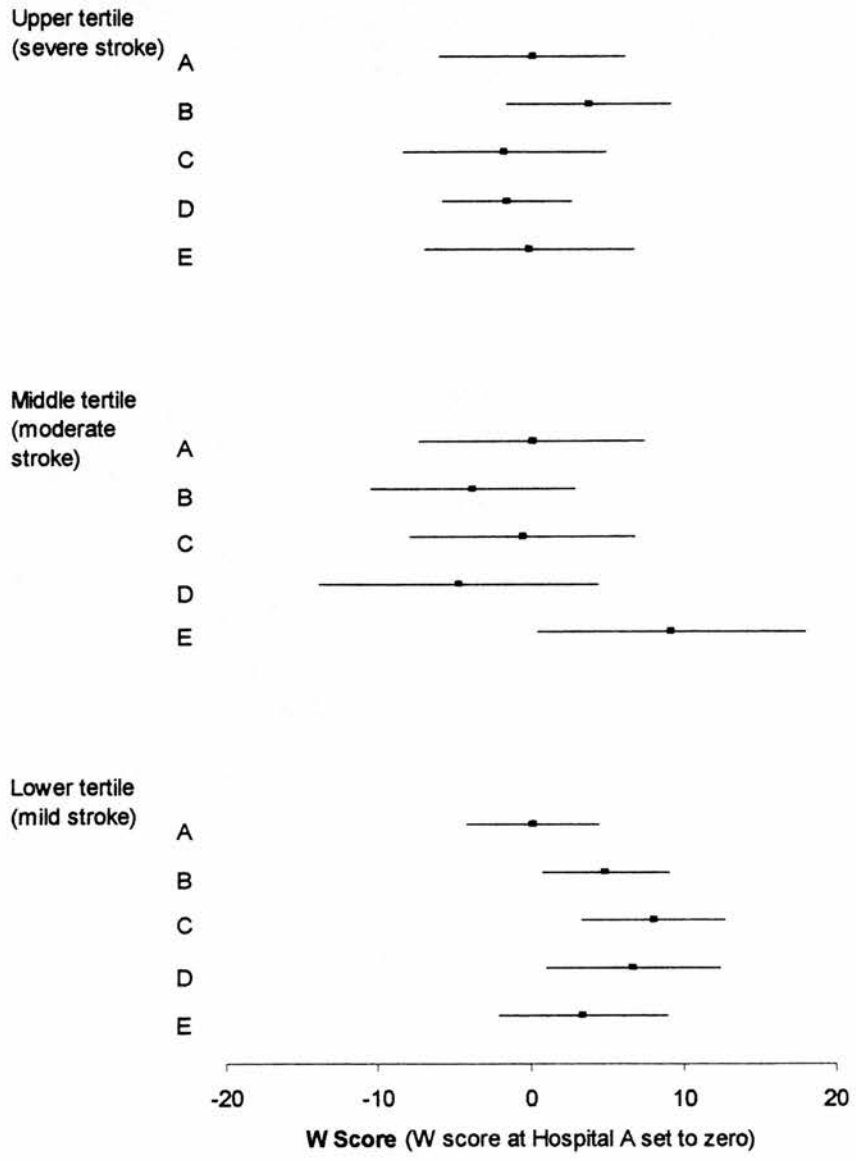
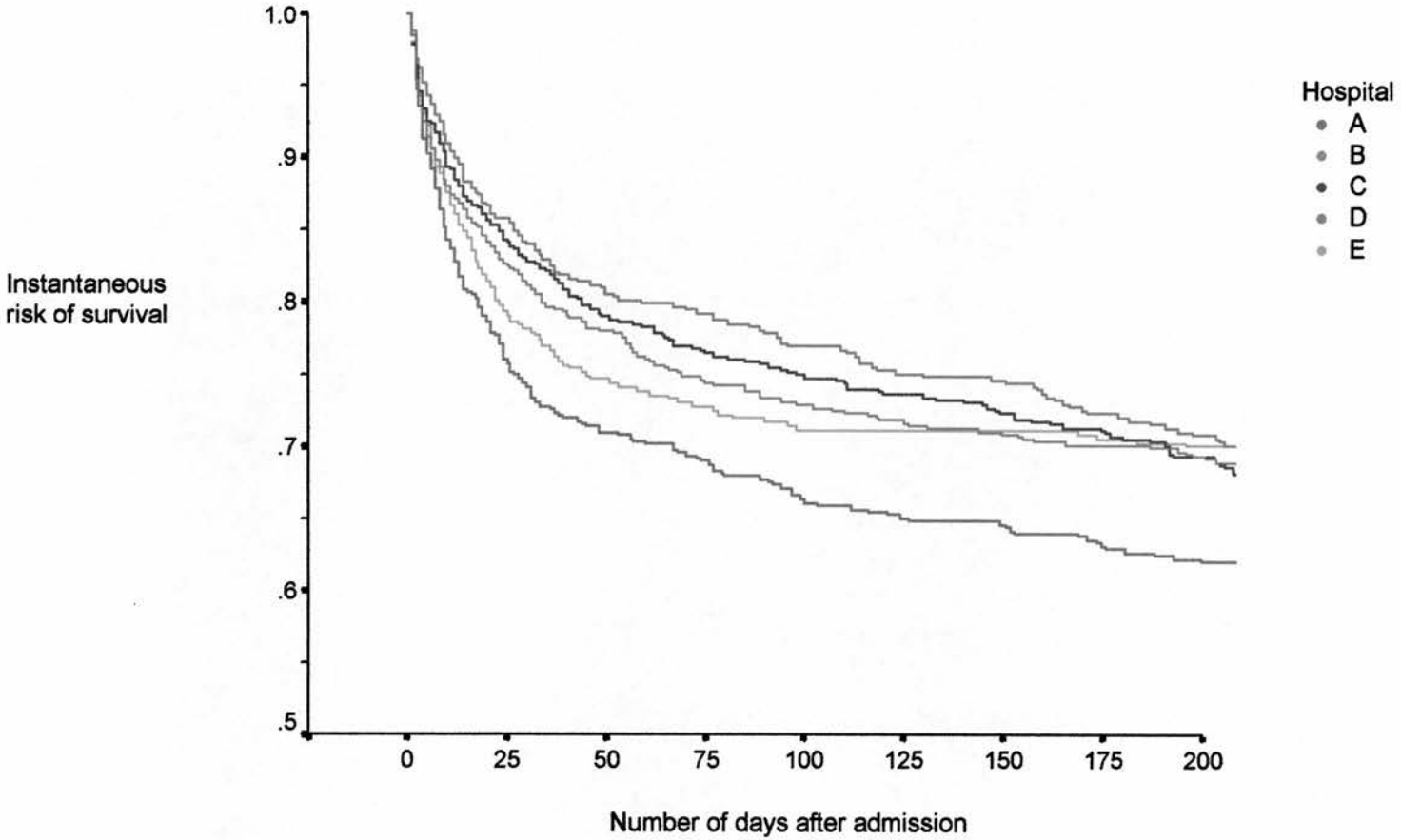


Figure 7.10 Adjusted survival curves (the instantaneous risk of survival after adjusting for casemix using the study model) per hospital; survival plotted till 200 days after admission.



Chapter Eight. SOP results (2): Structure and process of care and their link with outcomes after adjusting for casemix

8.1 Introduction

This chapter describes the results of the survey of the structure and process of stroke care at each hospital and relates them to the comparisons of outcomes, adjusted for casemix. The principal methods used to survey the structure and process of care have already been described in section 3.6.

8.2 Allowing for legitimate variation in the process of care

Our measurements of the process of care record only whether or not an item was provided. As such, they fail to take account of situations where the failure to provide particular items is appropriate, for example, not holding a MDT meeting for a patient discharged within a few days or not prescribing an anti-thrombotic drug after a haemorrhagic stroke. Simple comparisons of these data are likely, therefore, to be confounded by variation in casemix. I used two methods to minimise this bias. First, I stratified the comparisons of the process of care by the predicted risk of death at six months. Thus, I divided each hospital cohort into tertiles of predicted risk (see section 7.2.2) approximating to mild, moderate and severe stroke in order to make comparisons between groups with roughly similar treatment needs. I used box and whisker plots to demonstrate the predicted 'risk profile' of the hospital cohorts in order to gauge the success of this strategy. Second, where appropriate, I used other

data to refine the comparisons, for example, rather than compare the proportion of patients prescribed an anti-thrombotic drug I limited the comparison to those not shown to have had a haemorrhagic stroke and who were discharged alive; on the assumption that early hydration may reflect better quality of care (Indredavik *et al.* 1999), I limited the comparison of the use of parenteral fluids to the first 24 hours of admission; etc.

8.3 Results

8.3.1 The predicted prognoses of the hospital cohorts within each stratum

The 'risk profiles' of the hospital cohorts within each tertile of predicted risk of death at six months are shown in Figure 8.1. Within each box-plot the line bisecting the box indicates the median predicted risk, the ends of the box indicate the 25^{th} and 75^{th} percentile risks, and the whiskers indicate the 2.5^{th} and 97.5^{th} percentile risks. Within the middle and upper tertiles the predicted prognoses of the hospital cohorts were very similar. Within the lower tertile the predicted prognoses of the cohorts were also very similar, except at Hospital D, whose cohort had a modestly better prognosis.

8.3.2 Comparisons of the structure and process of care

The structures and processes of stroke care of each hospital are shown in Tables 8.1 to 8.4 and Tables 8.5 to 8.14, respectively. I have summarised the key findings below.

Access to specialised and organised stroke care

The provision of specialised and organised stroke care varied considerably between hospitals. The most important differences were between Hospital A and Hospitals B to E. Hospital A did not have a physician specifically responsible for its stroke services, did not provide any specialised rehabilitation for patients with stroke and only 10% of its patients had the results of a MDT meeting noted in their medical record, and only then after a median delay of 50 days. Hospitals B to E, on the other hand, each provided organised and specialised stroke care at some point during the study. However, access to organised services varied between these hospitals. Thus, in comparison with Hospitals B & C, double the proportion of patients were admitted to a SRU at Hospital E and two and a half fold more were discharged from the care of a specialist stroke physician at Hospital D. Only Hospitals B & D provided organised stroke care for patients not admitted to their SRU, the former by means of a mobile stroke rehabilitation team, the latter by informal extension of the SRU care to a nearby ward (as evidenced by the high proportion of patients who were under the care of a specialist stroke physician and who had a MDT meeting documented despite not being admitted to the SRU itself).

Access to specialised care also varied in subtle ways. The SRUs at Hospitals B & C and, for a few months, at Hospital E operated age related admission policies. Hospital B delayed admission to its SRU for an average of two or more weeks while the other hospitals admitted patients within a few days (hence the SRU at Hospital B admitted proportionately fewer patients with mild stroke and its median length of stay was longer). At Hospital C, patients deemed suitable for SRU care were admitted very quickly but access for patients with severe stroke was relatively poor (only a half to a third of that elsewhere). At Hospitals C & E, the SRUs opened half and three quarters of the way through the study, respectively. Once open, access to a SRU was in fact considerably greater than at Hospitals B & D.

Factors indicating the extent to which care truly was specialised and organised

The degree to which care was truly organised and specialised varied between hospitals. At Hospitals B & D, organisation was well established. The specialist stroke physicians were active members of the stroke interest and research community and the hospitals had written protocols describing assessment and treatment strategies, ran hospital stroke registers and provided specialist clinics for minor stroke and TIA. At Hospital D, virtually all patients admitted to the SRU came under the care of a specialist stroke physician and had a MDT meeting documented. The interval between admission and the first recorded MDT meeting was shortest at Hospital D and its SRU also stood out in terms of education of staff, for reporting the longest MDT meetings in relation unit size, for the regular attendance of nurses at therapy sessions and for running a carer support group throughout the study. At Hospital B, on the other hand, only about 60% of patients admitted to the SRU came

under the care of a specialist stroke physician or had a MDT meeting documented and the mobile stroke rehabilitation team functioned without any regular medical input.

At Hospitals C & E, organised stroke care developed during the study and it is likely that staff will have taken time to become experienced in working as an effective team. These hospitals did not have written protocols describing assessment and treatment strategies, did not run specialist clinics for minor stroke and TIA, and both appeared to allocate less time to their MDT meetings. At Hospital C the specialist stroke physician was not a member of any stroke interest group and, although MDT meetings were reported to be held regularly (according to our survey) the results were infrequently noted in the medical record. The delay between admission and the first recording of a MDT meeting was also much longer at Hospital C than elsewhere. At Hospital E, only about 40% of patients admitted to the SRU came under the care of the specialist stroke physician, the lowest proportion of any SRU, despite the fact that Hospital E admitted the lowest number of stroke patients overall. Also, the MDT at Hospital E did not appear to set formal rehabilitation goals and the nursing staff appeared disengaged.

Access to diagnostic imaging

Access to CT head scanning varied considerably. Hospital A scanned by far the smallest proportion of patients, particularly those with moderate and severe stroke, reflecting the fact that it was the only hospital without a CT scanner. Surprisingly,

Hospital B scanned the second lowest proportion and with the greatest delay. Hospital D scanned the highest proportion and with the least delay. All hospitals had access to carotid doppler ultrasound, echocardiography and video-fluoroscopy of the ability to swallow, although at Hospital A the latter required a 30 mile transfer.

The use of anti-thrombotic agents for secondary prevention

The use of anti-thrombotic agents in patients discharged alive and in whom a haemorrhage was not identified varied significantly. Their use was greatest at Hospital D and lowest at Hospital A. The overall variation in the use of anti-thrombotic drugs was principally driven by different prescribing practices in patients with severe stroke: these patients were much less likely to receive anti-thrombotic drugs at Hospitals A,B & C than at Hospitals D & E. When patients with ischaemic stroke only were considered the variation followed a similar but less marked pattern.

The Royal College of Physicians Stroke Audit Package

The median compliance with the RCPSAP criteria was highest at Hospital D, intermediate and similar at Hospitals B, C & E and lowest at Hospital A. Hospital A also had the lowest score and Hospital D the highest score for the individual criteria which measured whether a swallow assessment and a clear diagnostic formulation had been recorded within 24 hours of admission. Despite providing organised stroke care, Hospital B also scored relatively poorly on these items.

Non-evidence based medical interventions

The use of medical interventions without an accepted evidence base (at the time of the SOP) varied significantly. Hospital E gave parenteral fluids on the day of admission to the greatest proportion of patients while Hospital D gave the least. Hospital A inserted the fewest urinary catheters into patients with moderate and severe stroke (even in patients with severe stroke still alive after seven days i.e. those most 'at risk'). Almost no patients were prescribed subcutaneous heparin within 48 hours of admission at Hospital A while nearly 7% were at Hospital D.

Nursing staff and the Professions Allied to Medicine

The number of nursing staff per 100 beds to which stroke patients were usually admitted was similar across hospitals except Hospital D. Here the provision of nurses on its SRU was considerably higher than that on SRUs elsewhere but, surprisingly, the provision of nursing staff on its geriatric wards was considerably lower than that on geriatric wards elsewhere.

The number of therapists per 100 beds to which stroke patients were usually admitted varied between hospitals and especially between SRUs. The SRU at Hospital D had a consistently high provision of physiotherapists and occupational therapists in comparison with SRUs elsewhere while the provision of occupational therapists on the SRUs at Hospitals B & E was comparatively low. The provision of occupational therapists on the geriatric wards at Hospital A (where the majority of its rehabilitation took place) was also low. It should be noted that the provision of

therapists outside of the SRU at Hospital B is likely to have been higher than that shown because the calculation of therapists per 100 beds included the members of its mobile stroke team who, in reality, will have focused their activities on patients with stroke alone.

8.3.3 Ranking of hospitals by the overall quality of stroke care

Hospital A

The overall quality of care appeared to be lowest at Hospital A. To summarise, it was the only hospital which failed to provide any organised and specialised stroke care, which had the lowest access to brain imaging, which was the least likely to prescribe anti-thrombotic drugs for secondary prevention, which was the least likely to record a swallow assessment or diagnostic formulation on admission and whose medical records were, by some margin, the least complete in relation to the RCPSAP.

Hospital D

The overall quality of care appeared to be highest at Hospital D. Specifically, it provided the greatest access to specialist stroke physician care and recorded the greatest proportion of MDT meetings; it admitted the second greatest proportion of patients to a SRU *but* this care was more likely to be genuinely organised and specialised, and better staffed, than at Hospital E (which admitted the greatest proportion to a SRU). Hospital D also provided the best access to brain imaging,

was the most likely to prescribe anti-thrombotic drugs for secondary prevention, was the most likely to record a swallow assessment and clear diagnostic formulation on admission and had the most complete medical records in relation to the RCPSAP.

Hospitals B,C & E

The quality of care at Hospitals B,C & E was intermediate to that at Hospitals A & D (although generally closer to that at Hospital D than at Hospital A) and there was no consistent trend for one hospital to 'out perform' the others. In these respects, the overall quality of care at Hospitals B,C & E might be considered similar. However, there were a number of important differences. The organised service at Hospital B was mature and catered for patients on both the general wards and the SRU whilst at Hospitals C & E organised care was limited to the SRUs and, especially at Hospital E, appeared less specialised. On the other hand, Hospital B was the slowest to admit patients to a SRU, provided the worst access to brain imaging, was the least likely to record a swallow assessment and had the least complete medical records in relation to the RCPSAP. The capacity of the SRU at Hospital B was also much less than that at Hospitals C & E. Hospital E gave anti-thrombotic drugs to a greater proportion without a haemorrhagic stroke than Hospital C and parenteral fluids to a greater proportion with abnormal GCS than Hospitals B or C.

8.4 Discussion

8.4.1 The validity of the data describing the structure and process of stroke care

Before considering the relationship between our data describing the structure and process of care and the hospital outcomes it is important to consider the validity of our data and their suitability to the task of correlating care with its outcome. Regrettably, the data suffer from a number of shortcomings.

The validity of our data was compromised by our methods of measurement and by our partial knowledge of what constitutes effective stroke care. Our methods of measurement are likely to have been affected by two major biases. First, recording bias, the possibility that the data laid down in the medical record may not properly reflect the quality of the care that was provided. This bias is likely to be especially relevant to the RCPSAP criteria, to our identification of MDT meetings and to our non-evidence based items of care, such as the use of parenteral fluids, urinary catheters, etc. Fortunately, for the other aspects of care (the type of ward, the consultant in charge, CT scan date and result, the prescription of an anti-thrombotic drug) any recording bias is likely to be small because the items represent fundamental elements that are likely to be recorded in nearly all circumstances. The potential for recording bias may also have been greater at Hospital A due to its failure to use a stroke clerking proforma and its lack of a specialist stroke physician, factors which are known (Davenport *et al.* 1995b) or seem likely to increase the recording of key items. In defence of these shortcomings, it is arguable that in the context of stroke care, the completeness of the medical record is not so much a bias

but a subtle marker of quality since stroke care is most effective when delivered by an integrated team within which excellent communication and hence detailed and complete records are vital. There is also weak evidence from other fields of medicine that better recording is generally associated with better quality of care (Lyons and Payne 1974; Starfield *et al.* 1979; Zuckerman *et al.* 1975). However, whilst there is certainly something in these arguments, they are not sufficiently strong to negate the possibility of recording bias altogether.

The second potential reason for inaccurate measurement is observer error. As with our casemix data, this may have occurred because our audit assistant was aware of the hospital to which the patient was admitted and, in many cases, also their outcome (Caplan *et al.* 1991b; Gjorup *et al.* 1986). Again this is most likely to be relevant to the RCPSAP, MDT and non-evidence based criteria since the other criteria are less open to interpretation; and also to Hospital A since a greater proportion of its medical records were jumbled or incomplete (see section 7.4). Observer error, in the form of biased reporting, may also have affected the collection of data describing the structure of stroke care by interview, especially the more detailed items such as the average length of MDT meetings or the attendance of nurses at therapy sessions, since these are open to interpretation and even manipulation. However, our interviews involved several members of staff, minimising the impact of any individual's inaccurate report.

Our partial knowledge of what constitutes effective stroke care limited the validity of our structure and process data in two ways. Our limited ability to specify the essential elements of organised care (see section 3.6 and (Dennis 2000)) meant that it was difficult to set anything other than basic standards by which to judge performance. Thus, our main measurements of organised care were based only upon the proportions admitted to a SRU, the proportion with a recorded MDT meeting and the proportion discharged from the care of a specialist stroke physician. Comparison of these data between hospitals rests on the assumption that all SRUs, MDTs and stroke physicians are equally effective, and clearly this is unlikely to be the case. In mitigation, our structural survey and cross-tabulation of items allowed us to flesh out these measurements and provided some insight into the accessibility and working practices of each SRU. Nonetheless, our measurements remain somewhat superficial and certainly do not tell us how well the therapeutic activities of the team members were performed or how well they were co-ordinated toward appropriately selected rehabilitation goals. Indeed, this point highlights a wider failing, namely our failure to measure the very stuff of stroke care (the *therapeutic* activities of physicians, nurses and therapists) wherever it was provided, whether on a SRU or an ordinary ward. This failing was largely unavoidable. In the mid-1990s, there were almost no evidence or consensus based criteria by which to judge the value of these activities (a situation that, fortunately, is beginning to improve – see section 2.2) and, even if there were, many are either not routinely recorded e.g. the amount of time spent with a therapist; positioning and lifting techniques on the ward; effective early mobilisation; efforts to maintain morale; etc., or are of uncertain validity when retrospectively collected e.g. the identification and treatment of complications by

medical staff (Davenport *et al.* 1996a). As a result, for the majority of patients at each hospital (who were not admitted to a SRU) the content validity of our measurements of the process of care is low.

The other short-comings of our structure and process data relate to their suitability to the investigation of the relationship between the quality of care and its outcome. Here, McAuliffe (1978) suggests that the measured items of care should be: evidence based; emphasise therapeutic aspects (rather than just the recording of baseline data); influence the outcomes under investigation; make allowance for legitimate variation; and be combined in such a way that the relationship between the *overall* quality of care and its outcome is measured. Our measurements do indeed emphasise the provision of some items of care that are proven or strongly held to influence the study outcomes and the items that are not directly therapeutic are important and are likely to facilitate the provision of items that are. However, to varying degrees, our data fall short of the remainder of McAuliffe's criteria.

The most important of these shortcomings is the already mentioned failure to *directly* measure the therapeutic activities of physicians, nurses and therapists in all cases. These activities are the very ones that have an impact on death and disability, the outcomes under investigation. Since the majority of patients at each hospital were not admitted to a SRU, it follows that we have failed to directly measure those aspects of care which impact the study outcomes in the majority of patients. As a

result, our ability to consider whether any overall differences in death or dependency might have been *caused* by the observed differences in care is limited.

Our data did make allowance for legitimate variation in provision of care using time limits and extenuating circumstances for the RCPSAP criteria (see 3.6.2) and stratification by predicted prognosis for the other criteria. Stratification is likely to have been moderately successful: the cohorts within each stratum had remarkably similar prognoses and so, in broad terms, similar treatment needs. Also, our principal criteria are largely applicable irrespective of casemix e.g. it is arguable that nearly all patients ought to be admitted to a SRU or have a CT scan, etc. However, neither method, but especially stratification, can have accounted for the many and subtle factors that influence the provision of care. Furthermore, by shrinking the size of the cohorts, stratification reduced the power of the comparisons. Hence, it remains possible that even moderate sized differences in the provision of care between hospitals may still reflect differences in casemix and/or the play of chance.

Lastly, I did not construct an objective composite measure (a quality index) to quantify the *overall* quality of care at each hospital. I considered that the validity of a quality index would be open to considerable doubt given its need to measure 'apples and oranges' using the same, arbitrary yardstick and its inability to account for the inter-linked nature of stroke interventions. Instead, I provided a quasi-objective summary of the *overall* quality of care, leaving the informed reader to agree or disagree. The drawbacks to this method are its subjectivity, its imprecision

(reducing the power to differentiate between hospitals) and the inability to formally correlate quality of care with outcome.

In summary, our data describing the structure and processes of stroke care are not ideal. In particular, it is possible that their validity may be lower at Hospital A than at the other four hospitals. Our limited measurements and the format of the data also restricts the extent to which any correlation between quality of care and its outcome can be tested. Nonetheless, it is important to understand that many of these shortcomings were fundamentally insurmountable because of our basic lack of knowledge about stroke care (especially as it was in the mid 1990s) and because of the practical impossibility of prospectively measuring the numerous and on-going activities of routine care across several hospitals. As such, the quality of our data is, in all probability, about the best that could have been achieved under the circumstances of a large, retrospective and simple study and it should be borne in mind that they still provide a reasonable description of important aspects of stroke care. Indeed, perhaps the most important observation is that the collection of a relatively few items describing the structure and process of care appear able to identify a range of opportunities to improve the quality of stroke care at most if not all hospitals. It is also notable that many of these structure and process items are very simple and might lend themselves to routine data collection.

8.4.2 Relationship between the quality of the structure and processes of stroke care and its outcome, after adjusting for casemix

The survey of structure and process of care suggests that, overall, the quality of care for patients with stroke was lowest at Hospital A, highest at Hospital D and intermediate at Hospitals B,C & E. It will be recalled that, after adjusting for casemix, the hospital outcomes data mirrored this result: Hospital A had the worst outcome (in terms of case fatality), Hospital D had the best outcome (in terms of death or dependency) and Hospitals B,C & E had intermediate and largely indistinguishable outcomes. On first inspection, therefore, it would appear that there is indeed a relationship between routinely collected stroke outcomes (adjusted for casemix) and the quality of stroke care, in turn raising the possibility that a routine system of measuring stroke outcomes, based on the methods described in this study, might well act as a valid and practical indicator of the quality of stroke care. However, the study findings are more complex than they first appear.

The first considerations are practical. In the real world, it is unlikely that the routine measurement of dependency at a defined time after discharge will be attempted in the near future. For the medium term, it is likely that routine systems will have to use case fatality data alone or, as is currently the case in the UK, a combination of case fatality and place of residence data (measured as destination at time of discharge) as the key measures of outcome after stroke. Our findings suggest that, even under the relatively ideal circumstances of the SOP, these systems are likely to be of only limited value. Specifically, it appears that the systems might have a role in

identifying hospitals with *marked* shortcomings in stroke care. Thus, Hospital A had the highest adjusted case fatality and, by some margin, the lowest quality of stroke care. However, the same systems would appear to be insensitive to more moderate yet clearly important differences in care, such as the two fold variation in the proportion admitted to a SRU or the two and a half fold variation in the proportion cared for by a specialist stroke physician noted between Hospitals B to E. Indeed, it is notable that, on its own, the outcome of alive & at home was insensitive to *any* of the variations in the quality of care in the SOP.

These findings are important for two reasons. First, Hospitals B to E are likely to be representative of many, perhaps even the majority, of hospitals in Scotland which provide average to good stroke services but which have scope to improve in one or other aspect of care. A system of measuring the quality of stroke care based upon the outcomes of case fatality and residence data would clearly fail to identify these opportunities. Worse, by failing to flag them up, the system would run the risk of engendering an attitude of complacency rather than one of critical self reflection i.e. it might actually be detrimental to attempts to improve the quality of stroke care at many hospitals. Second, our findings raise the possibility that dichotomised residence data (institutional care vs. other) may be peculiarly unsuited to the task of measuring the quality of stroke care. Three reasons are likely to underlie this observation: first, the simple dichotomisation of residence data may obscure real but subtle differences in residential outcome (see section 7.7); second, because it so often equates to marked dependency, long term institutional care may be peculiarly determined by initial stroke severity and relatively little by hospital care; and third,

long-term institutional care is heavily dependent on local cultural factors (the wishes and expectations of patient and family) and social resources (whether the patient lived alone before the stroke, the size and layout of their house, the availability of community support, etc.) over which the hospital has no control (Brosseau *et al.* 1996; McKenna *et al.* 2002). Rudd et al (2001a) noted similar difficulties when trying to interpret variations between UK regions in the proportions of patients discharged to institutional care, the considerable differences not being explicable in terms of pre-stroke accommodation or disability, age, gender, length of stay (a proxy for stroke severity), the type of unit in which the majority of the stay took place or disability at discharge. Taken together, these observations underline the need for great caution when interpreting comparisons of rates of institutionalisation between hospitals, the need for further study of the reasons for the residual variation and even the need to reconsider the routine publication of this outcome.

In time, however, it may become possible to routinely measure dependency after stroke and so to report the combined outcome of 'death or dependency', the most clinically relevant outcome measured by the SOP. As a more rounded measure of outcome, one might expect death or dependency to act as a more sensitive indicator of the quality of stroke care. Somewhat surprisingly, this was not the case in the SOP. Using it, we were able to identify that Hospital D had the best quality of care but entirely failed to identify that Hospital A had the worst. The indications of quality of care based on death or dependency were also quite different to the indications based on case fatality data. Indeed, the correct identification of the hospitals with the best, intermediate and worst quality of quality of care was only

possible when *both* case fatality and 'death or dependency' were reported. The finding of non-uniform performance across different measures of outcome has been noted previously in other fields of medicine (Hartz.A.J. and Kuhn 1994; Iezzoni *et al.* 1994; Silber *et al.* 1995; Silber and Rosenbaum 1997). The suggested explanation for the phenomenon is that different outcomes are influenced by different aspects of care, the quality of which may vary within a given hospital. For stroke, it is arguable that case fatality is primarily influenced by the quality of management in the first few weeks (the majority of deaths after stroke occur within 30 days), in particular the prevention and treatment of the complications of immobility and cardiovascular disease; and that dependency is strongly influenced by the quality of the rehabilitation services (Stroke Unit Trialists' Collaboration 1997b). Following this logic, it is perhaps not a surprise that the optimum measurement of the quality of stroke care using outcomes seems to require the measurement of case fatality and functional dependency.

However, before we adopt adjusted case fatality and death or dependency data as legitimate indicators of the quality of stroke care, two other aspects of our findings merit consideration. First, it is important to note that we have studied only a small number of hospitals and so have provided only one example where each outcome correctly identified a hospital with better or worse stroke care. Our findings therefore do not tell us anything about the accuracy (the sensitivity and specificity) with which the proposed system of outcome measurement would be able to differentiate hospitals with exceptionally good or bad care from those with acceptable standards of care. As noted (see section 1.5.4), unless sensitivity and

specificity are high, the system would be of limited value and would also be harmful in terms of wasted resources and needlessly dented reputations. Regrettably, because a proper determination of sensitivity and specificity will require the study of tens of hospitals, for reasons of cost, it is uncertain that it will ever take place.

Second, it is also important that we reconsider the credibility of our findings; in particular, to question whether the residual differences in outcome can plausibly have *resulted* from the observed differences in care. If this were the case, two hypotheses follow. At Hospital A one should find deficiencies in the initial management of stroke relative to Hospitals B to E but rehabilitation of about equal quality to that at Hospitals B,C & E; and at Hospital D one should find initial management of equal quality to that at Hospitals B,C & E but rehabilitation of a much higher quality. As noted, the relative crudeness of our structure and process of care data limits our ability to explore these hypotheses. Nonetheless, various lines of evidence can be followed.

First, one may consider whether the differences in outcome might have been caused by the measured differences in the provision of organised rehabilitation, by far the most efficacious intervention for stroke that was routinely available in the mid 1990s. To illustrate this, consider Hospitals A & D. In comparison to conventional care, it is estimated that 22 patients (95% CI 10 to ∞) need to be treated on a SRU to prevent one death and 16 (95% CI 10 to 25) to prevent one case of death or dependency (Stroke Unit Trialists' Collaboration. 1997a). Since Hospital D admitted 154 patients

to a SRU and Hospital A none, it is reasonable to assume that this difference led to about seven fewer deaths and ten fewer cases of death or dependency at Hospital D, equating to 1.3 fewer deaths and 1.8 fewer cases of death or dependency per 100 admissions. After adjusting for casemix, there were in fact 5.6 fewer deaths and 11.3 fewer cases of death or dependency per 100 admissions at Hospital D i.e. the measured difference in SRU care is likely to explain only a fraction of the residual difference in each outcome between the two hospitals. Even if one assumes that equally good organised care was extended to all 292 patients with a recorded MDT at Hospital D, this would still have led to only 2.4 fewer deaths and 3.3 fewer cases of death or dependency per 100 admissions i.e. it would still explain only about half of the residual difference in case fatality and about a quarter of the residual difference in death or dependency. Given that the differences in the provision of organised care between the remaining hospitals were of a similar size or smaller, it is even less likely that they can explain their equally large residual differences in case fatality or death or dependency.

Thus, if they are the result of differences in the quality of care, the majority of the residual differences in outcome must be due to differences in therapeutic aspects of care that we did not measure. This possibility can only be examined by inference. Three observations support the hypothesis at Hospital A i.e. that Hospital A provided the worst initial management of stroke but rehabilitation of a similar quality to that at Hospitals B,C & E. First, Hospital A had the worst performance in terms of several mainly non-therapeutic markers of initial management (the completeness of the initial history and examination, the recording of a swallow assessment, the recoding

of a clear diagnostic formulation and the provision of CT head scan); second, the higher adjusted case fatality at Hospital A was mainly the result of higher case fatality in patients with severe stroke i.e. just those in whom shortcomings in the prevention and treatment of complications and cardiovascular disease would be most likely to lead to death; and third, the split nature of Hospital A provides a plausible mechanism. All patients at Hospital A were admitted to the main hospital but only transferred to the affiliated hospitals if they required on-going rehabilitation. It is possible that the main hospital may have provided poor initial care whilst the affiliated hospitals, each with a specialist geriatric rehabilitation ward, may have provided rehabilitation of a similar quality to that at Hospitals B,C & E. There are some inconsistencies with these lines of reasoning, in particular, Hospital A was not deficient in some markers of initial management and there was little documented evidence of co-ordinated team working on the geriatric rehabilitation units. Nonetheless, a plausible argument can be made that the outcomes at Hospital A might have been caused its methods of care.

The situation in relation to Hospital D is somewhat different. To explain its outcomes, it should have provided initial care of equal quality to Hospitals B,C & E but rehabilitation of a much higher standard. The structure of stroke care at Hospital D suggests a possible mechanism. Here, the majority of patients were admitted to the general wards (where initial care may have been of 'average' quality) but about half subsequently came under the care of the specialist stroke physician (who may have focused excellent organised rehabilitation services on those most likely to benefit). However, this argument is difficult to sustain. First, some of our

measurements of initial management at Hospital D suggest that it was, in fact, somewhat better than that at Hospitals B,C & E. Thus, Hospital D had the greatest proportion with a recorded swallow assessment and diagnostic formulation, the greatest and fastest access to CT head scanning, the least use of indwelling urinary catheters and the greatest proportion prescribed subcutaneous heparin (reflecting greatest involvement in the International Stroke Trial (International Stroke Trial Collaborative Group 1997a)). Second, only a small proportion of the lower adjusted death or dependency at Hospital D is explicable by its differential provision of organised stroke care (see above). Thus, the *great majority* of its lower level of dependency must have arisen from superior rehabilitation provided to patients who remained on the general wards, who were not under the care of the stroke physician and who, in the same setting, apparently received initial care of only average quality. This is difficult to believe. Third, and most important, the residual differences in death or dependency between Hospital D and Hospitals B,C & E were very large and occurred overwhelmingly in patients with the least severe stroke (see section 7.6.6), which is to say, in patients at relatively low risk of being dead or dependent at six months regardless of intervention. Hence, it is hard to conceive that any differences in care can have caused the large differences in dependency that were found. Thus, it is difficult to argue that that the outcomes at Hospital D are likely to have resulted from its methods of care.

The final consideration is the quality of the data describing structure, process and outcome on which all these deliberations are based. As detailed in chapter seven and here, they clearly suffer from a host of actual and potential shortcomings which,

together, might plausibly explain much or even all of the observed associations between better adjusted outcome and better quality of care. In particular, rather than worse quality of care, a reasonable alternative explanation of the higher adjusted case fatality at Hospital A may simply be systematic underestimation of baseline risk together with the play of chance. Similarly, the finding that Hospital A had the 'worst' quality of care may also, in part, reflect biased measurement. Given the difficulty in ascribing the very much lower level of death or dependency at Hospital D to better care, a plausible alternative explanation may simply be inadequate adjustment for the very much more favourable baseline characteristics of its patients.

Thus, the findings of our investigation into the relationship between routinely collected and adjusted outcomes and the quality of stroke care are complex and somewhat inconclusive. In many respects, our findings appear to support the use of case fatality and death or dependency data as valid indicators of the quality of stroke care. However, this optimistic conclusion should be treated cautiously. The sensitivity and specificity of the method has obviously not been established and detailed consideration of our findings casts a degree of doubt on the causal nature of the observed associations between better adjusted outcome and better quality of care. Even if our findings are valid, in the absence of dependency data, a routine system of quality measurement based on case fatality and institutionalisation (i.e. an enhanced version of systems currently in place in the UK) might only be sensitive to the most marked deficiencies in care; indeed, it appears that institutionalisation data may be altogether unsuited to the routine measurement of the quality of stroke care. To all this can be added the observation that even the most sensitive outcome-based system

(the combination of case fatality and death or dependency data) failed to identify the important differences in quality of care between Hospitals B,C & E, the potential shortcomings at Hospital D (in terms of the prescription of parenteral fluids and the provision of nursing staff on its geriatric wards) or the apparently excellent aspects of initial management at Hospital A.

In summary, a key purpose of the SOP was to investigate the utility of an idealised yet routinely feasible outcomes-based system of measuring the quality of stroke care. In answer to this question, our findings suggest that, if ever implemented in full, our methods *might* indeed be useful in identifying hospitals with outlying standards of care but that if only partially implemented (i.e. if only case fatality and residence data are collected) the system *might* only be useful in identifying hospitals with *marked* shortcomings in care. Both systems, but especially the latter, will undoubtedly fail to identify important opportunities to improve the quality of stroke care at many, perhaps most, hospitals with seemingly average outcomes. These uncertain and somewhat disappointing conclusions derive from a method of measuring and comparing outcomes far superior to that which is currently in use in the UK. Given this state of affairs, one must ask oneself whether there is not a more certain and useful way that we can measure the quality of hospital stroke care. This possibility is considered in the final chapter of this thesis.

Summary

1. Important aspects of the structure and process of stroke care varied substantially between the five study hospitals. Overall, the quality of stroke services appeared highest at Hospital D, intermediate at Hospitals B,C&E and lowest at Hospital A.

2. This ranking matched the ranking based upon the combined outcomes of case fatality and death or dependency at six months, adjusted for casemix, suggesting that our system for measuring outcome *might* provide a valid method by which to routinely identify hospitals with particularly high or low quality of care.

3. A partial system that reported only case fatality at six months, adjusted for casemix using our methods, would be less sensitive but *might* have a limited role in identifying hospitals with very marked shortcomings in stroke care.

4. Despite its apparent face validity, institutionalisation appears to be an insensitive indicator of quality of stroke care. Routine comparisons of institutionalisation should be interpreted with particular caution and its use as a routinely published quality indicator reconsidered.

5. All these conclusions must remain tentative because of the small number of hospitals studied and the potential continuing impact of biased measurement, variation in casemix and the play of chance. The findings of the SOP should be regarded only as possibilities that merit further investigation.

6. Even our 'ideal' system would fail to identify moderate but potentially important opportunities to improve the quality of care at many if not most hospitals.

7. Simple structure and process data clearly provide a rich source of information with which to understand and plan improvements in stroke care. However, their interpretation is also problematic, in particular because of our still limited library of robust evidence linking structures and processes to outcome, the difficulty in generalising the efficacy data that we have, and our reliance on information ordinarily laid down in the medical record.

Table 8.1 The structure of stroke care: organisation and specialisation

	Hospital A	Hospital B	Hospital C	Hospital D	Hospital E
Physician specifically responsible for patients with stroke	No	Yes	Yes	Yes	Yes
Stroke interest groups to which the stroke physician belongs	-	BSRG, BASP, CHSS (council member)	None	BSRG, BASP, SIGN,CHSS	BSRG
Stroke rehabilitation unit (SRU)	No	Yes	Yes	Yes	Yes
Acute stroke unit	No	No	No	No	No
Acute patients also admitted to SRU	-	No	Yes	Occasionally	Yes
Mobile stroke team for patients not admitted to SRU	No	Yes	No	No	No
SRU location	-	Subsidiary site	Main site	Main site	Main site
Year opened	-	1993	Sep.1996	1993	Feb.1996
Proportion of study during which SRU operative	-	100%	46%	100%	71%
Number of SRU Beds	-	30	18	15	30
SRU beds per 100 stroke admissions per year*	-	8	7	5	16

Does the ward housing the SRU also admit non-stroke patients	-	Stroke only	6 additional beds for general rehab.	Stroke only	Variable additional beds for general rehab.
Ability to expand SRU when usual number of beds fully occupied	-	Yes, to nearby general rehab wards	Yes, within own ward	Yes to adjacent ward – patients integrated into SRU	Yes, within own ward
Physical setting of SRU	-	Purpose built ward	Converted ward	Converted ward	Converted ward – staff comment "cramped" setting
Hospital stroke register	No	Yes	No	Yes	Yes (not final 4 months)
Written protocols for assessment / treatment of stroke patients	No	Yes	No	Yes (in year 2)	No
Stroke clerking proforma available and used	No	SRU throughout GA wards intermittent	SRU throughout GA wards throughout GM wards 2nd year	SRU throughout	SRU throughout GA wards intermittent

BSRG British Stroke Research Group
BASP British Association of Stroke Physicians
CHSS Chest, Heart and Stroke Scotland
SIGN Scottish Inter-collegiate Guidelines Network
GA Geriatric medicine assessment ward
GM General medicine ward

* Annual number of patients admitted with stroke derived from SOP data

Table 8.2 The structure of stroke care: access to diagnostic facilities

	Hospital A	Hospital B	Hospital C	Hospital D	Hospital E
CT scanner	None on site	On site	On site	On site	On site
		24 hours	Not nights/ weekends	24 hours	Year1: not nights / weekend Year 2: 24 hours
	13 or 30 mile transfer for scans		24 mile transfer for 'out of hours' scan		27 mile transfer for 'out of hours' scan
MRI scanner	On site (last 3 months only)	None on site	None on site	On site	None on site
	Prior 30 mile transfer if required	4 mile transfer if required	24 mile transfer if required		20 mile transfer if required
Carotid ultrasound scanner	Yes	Yes	Yes	Yes	Yes
Echocardiography	Yes	Yes	Yes	Yes	Yes
Video-fluoroscopy of swallow	None on site	Yes	Yes	Yes	Yes
	30 mile transfer if required	Temporarily shut at year 1 survey			

Table 8.3 The structure of stroke care: nurses, physiotherapists and occupational therapists (all grades); number (whole time equivalents) per 100 beds to which stroke patients are usually admitted (as of May 1997)

	Hospital A	Hospital B	Hospital C	Hospital D	Hospital E
Nurses					
General medical wards	99	89	90	{94}	98
Geriatric wards	87	96	95	61	91
SRU	-	93	86	125	99
Physiotherapists					
General medical wards	1.7	2.7	2.1	1.6	1.8
Geriatric wards	5.1	3.9	7.7	4.6	4.3
SRU	-	8.3	8.3	15.3	11.0
Occupational therapists					
General medical wards	2.4	1.3	2.0	2.1	1.0
Geriatric wards	2.5	2.4	4.3	3.9	9.3
SRU	-	2.3	13.9	10.0	3.3

Table 8.4 The structure of stroke care: staff supporting discharge from hospital and outpatient facilities

	Hospital A	Hospital B	Hospital C	Hospital D	Hospital E
Stroke liaison nurse	No, but One of the subsidiary hospitals had a full-time outreach nurse who devoted 70% of his time to stroke patients	Yes For patients admitted to main site only (i.e. not patients on SRU) Part-time year 1 Full-time year 2	Yes (last 11 months only) All patients	Yes (first 10 months only) For patients admitted at main site only as part of clinical trial, contract terminated at end of trial	No
Social workers	Hospital generic & community	Hospital generic Stroke specific at SRU	Hospital generic & community	Year 1 Hospital generic Year 2 Main site none attached; Subsidiary site hospital generic	Hospital generic & community
Geriatric Medicine Day Hospital & Outpatients PAMS	Yes	Yes	Yes	Yes	Yes
Specialist clinic for the rapid diagnosis and management of minor stroke and TIA	No	Yes	No	Yes	No

Table 8.5 The process of stroke care: proportion of patients admitted to a Stroke Rehabilitation Unit (SRU)

	No. of cases overall	Hospital A %	Hospital A d	Hospital B %	Hospital B d	Hospital C %	Hospital C d	Hospital D %	Hospital D d	Hospital E %	Hospital E d	p (χ^2)
All patients	2714	**0**	520	**21**	744	**21**	520	**28**	545	**40**	385	<0.0001
By predicted risk of death at 6 months												
Lower tertile	905	**0**	130	**12**	225	**18**	147	**23**	280	**29**	123	<0.0001
Middle tertile	906	**0**	174	**32**	263	**33**	189	**44**	146	**54**	134	<0.0001
Upper tertile	903	**0**	216	**19**	256	**10**	184	**22**	122	**34**	128	<0.0001
Patients ≤ 65 years old	638	**0**	111	**4**	185	**0**	105	**24**	155	**21**	82	<0.0001
All patients in the second year of the study only	1380	**0**	271	**21**	352	**43**	254	**29**	308	**54**	195	<0.0001
Delay between hospital admission and transfer to SRU in days (median & IQR)	572	-	-	**18**	(13 to 27)	**1**	(1 to 5)	**3**	(1 to 9)	**3.5**	(1 to 8)	<0.0001KW
Length of stay on SRU in days (median & IQR)	572	-	-	**50**	(29 to 109)	**27**	(14 to 63)	**31**	(12 to 83)	**24**	(13 to 43)	<0.0001KW
Of patients admitted to SRU, proportion of stay spent on SRU (% & 95% CI)	572	-	-	**73**	(57 to 86)	**94**	(82 to 98)	**91**	(75 to 99)	**87**	(68 to 97)	<0.0001

KW Kruskal Wallis test

Table 8.6 The process of stroke care: selected aspects of usual practice within each SRU (data obtained by staff interview)

	Hospital B	Hospital C	Hospital D	Hospital E
Age related admission policy	> 60 years (but stroke team for ≤ 60 years)	> 65 years	All ages	65 years 1st year All ages 2nd year
Staff training & education				
Written material	General leaflets	General leaflets	Comprehensive pack	General leaflets
Educational meetings	Yes	Yes	Yes	Yes
Education for patient/carers				
Written material	Yes	Yes	Yes	Yes
Educational meetings	Yes	No	Yes	Last 6 months only
Rehabilitation				
Setting of rehabilitation goals	Yes	Yes	Yes	No?
Documentation of rehab. goals	Medical record (therapists share these)	Medical record (sometimes) + therapist records always	Medical record (sometimes) + all therapists records always	"Not documented as MDT do not consider they have time to discuss them"
Formal assessment of progress	FIM score weekly	Barthel score weekly	Barthel score weekly	Barthel score weekly
MDT meetings	Weekly	Weekly	Weekly	Weekly

MDT decisions noted in medical-record	Yes (therapists share these)	Infrequent	Yes	Yes
Usual duration of MDT meeting	120 mins	30 to 60 mins	120 mins	60 to 90 mins
Nurses attend MDT meeting	Yes	Yes	Yes	"Often too busy"
Nurses feel involved in providing rehabilitation	Yes	Yes	Yes	"Often too busy"
Nurses regularly attend therapy sessions with patient	Only if required	No	Yes	No
Discharge				
Pre-discharge home visit	Nearly all cases	Nearly all cases	Nearly all cases	Nearly all cases
Families involved in care in hospital	Yes	Yes	Yes	Yes
Carers support group	No	No	Yes	Last 6 months only

FIM Functional Independence Measure

Table 8.7 The process of stroke care: proportion of patients discharged from the care of a consultant specifically responsible for stroke

	No. of cases overall	Hospital A		Hospital B		Hospital C		Hospital D		Hospital E		
		%	d	%	d	%	d	%	d	%	d	p (χ^2)
All patients	2724	**0**	521	**21**	746	**23**	520	**48**	551	**30**	386	<0.0001
By predicted risk of death at 6 months												
Lower tertile	908	**0**	131	**20**	226	**14**	147	**44**	281	**31**	123	<0.0001
Middle tertile	908	**0**	174	**28**	263	**39**	189	**58**	148	**33**	134	<0.0001
Upper tertile	908	**0**	216	**16**	257	**15**	184	**45**	122	**26**	129	<0.0001
Patients admitted to hospital for over 7 days	2044	**0**	367	**26**	589	**27**	407	**58**	394	**34**	287	<0.0001
Patients admitted to SRU	572	-	-	**58**	158	**53**	108	**97**	154	**39**	152	<0.0001

Table 8.8 The process of stroke care: the proportion of patients in whom a Multi-Disciplinary Team (MDT) meeting was documented at any time in the medical record

	No. of cases overall	Hospital A %	Hospital A d	Hospital B %	Hospital B d	Hospital C %	Hospital C d	Hospital D %	Hospital D d	Hospital E %	Hospital E d	p (χ^2)
All patients	2705	**10**	518	**42**	743	**15**	519	**54**	541	**48**	384	<0.0001
By predicted risk of death at 6 months												
Lower tertile	902	**6**	130	**46**	225	**11**	147	**43**	277	**38**	123	<0.0001
Middle tertile	903	**13**	172	**52**	263	**26**	188	**77**	146	**61**	134	<0.0001
Upper tertile	900	**8**	216	**28**	255	**8**	184	**53**	118	**43**	127	<0.0001
Patients admitted to hospital for over 7 days	2028	**13**	365	**53**	586	**20**	406	**71**	385	**62**	286	<0.0001
Patients admitted to SRU	572	-	-	**61**	158	**46**	108	**94**	154	**86**	152	<0.0001
Delay between hospital admission and first recorded MDT meeting in days (median & IQR)	918	**52**	(29 to 70)	**8**	(5 to 14)	**21**	(11 to 35)	**6**	(3 to 11)	**8**	(5 to 13)	<0.0001KW

KW Kruskal Wallis test

Table 8.9 The process of stroke care: proportion of patients with a CT or MRI head scan performed

	No. of cases overall	Hospital A %	Hospital A d	Hospital B %	Hospital B d	Hospital C %	Hospital C d	Hospital D %	Hospital D d	Hospital E %	Hospital E d	p (χ^2)
All patients	2712	**53**	520	**76**	744	**81**	520	**87**	544	**83**	384	<0.0001
By predicted risk of death at 6 months												
Lower tertile	905	**86**	130	**88**	225	**94**	147	**95**	280	**96**	123	0.001
Middle tertile	906	**68**	174	**89**	263	**92**	189	**92**	146	**91**	134	<0.0001
Upper tertile	901	**22**	216	**53**	256	**59**	184	**62**	118	**61**	127	<0.0001
Patients discharged alive & prescribed an anti-thrombotic drug	1369	**80**	207	**92**	382	**94**	245	**95**	344	**96**	191	<0.0001
Of patients with CT scan, proportion scanned within 7 days of admission	2046	**68**	276	**55**	562	**92**	419	**92**	472	**92**	317	<0.0001
Delay between hospital admission and head scan in days (median & IQR)	2046	**5**	(2 to 8)	**5**	(2 to 10)	**1**	(1 to 3)	**1**	(0 to 2)	**2**	(1 to 4)	<0.0001KW

KW Kruskal Wallis test

Table 8.10 The process of stroke care: proportion of patients discharged alive who were prescribed an anti-thrombotic medication

	No. of cases overall	Hospital A %	Hospital A d	Hospital B %	Hospital B d	Hospital C %	Hospital C d	Hospital D %	Hospital D d	Hospital E %	Hospital E d	p (χ^2)
Patients without an identified cerebral haemorrhage (infarcts + type unknown)	1665	**73**	282	**82**	465	**76**	322	**90**	376	**86**	220	<0.0001
By predicted risk of death at 6 months												
Lower tertile	764	**88**	112	**86**	198	**83**	116	**93**	239	**93**	99	0.033
Middle tertile	637	**70**	117	**86**	195	**77**	138	**88**	103	**83**	84	0.001
Upper tertile	264	**51**	53	**58**	72	**60**	68	**82**	34	**76**	37	0.016
Patients with ischaemic stroke only	1468	**81**	204	**85**	412	**78**	290	**91**	356	**88**	206	<0.0001

Table 8.11 The process of stroke care: proportion of patients given parenteral fluids on the day of admission to hospital

	No. of cases overall	Hospital A		Hospital B		Hospital C		Hospital D		Hospital E		
		%	d	%	d	%	d	%	d	%	d	p (χ^2)
All patients	2701	**44**	518	**38**	740	**42**	517	**31**	542	**46**	384	<0.0001
By predicted risk of death at 6 months												
Lower tertile	899	**18**	130	**14**	223	**23**	146	**18**	277	**25**	123	0.059
Middle tertile	902	**40**	173	**37**	262	**36**	187	**40**	146	**39**	134	0.89
Upper tertile	900	**62**	215	**60**	255	**64**	184	**52**	119	**72**	127	0.022
In patients with abnormal GCS Eye score (i.e. those unconscious/drowsy)	650	**67**	154	**66**	179	**67**	135	**62**	91	**75**	91	0.06

Table 8.12 The process of stroke care: proportion of patients in whom an indwelling urinary catheter was inserted

	No. of cases overall	Hospital A		Hospital B		Hospital C		Hospital D		Hospital E		
		%	d	%	d	%	d	%	d	%	d	p (χ^2)
All patients	2704	**36**	519	**43**	741	**45**	519	**32**	542	**43**	383	<0.0001
By predicted risk of death at 6 months												
Lower tertile	901	**12**	130	**13**	224	**14**	147	**11**	277	**16**	123	0.74
Middle tertile	903	**31**	173	**40**	263	**41**	188	**45**	146	**43**	133	0.102
Upper tertile	900	**54**	216	**74**	254	**72**	184	**62**	119	**71**	127	<0.0001
Upper tertile patients still alive 7 days after admission	420	**59**	100	**90**	117	**80**	88	**68**	57	**78**	58	<0.0001

Table 8.13 The process of stroke care: proportion of patients given subcutaneous heparin on the day or day after admission

	No. of cases overall	Hospital A		Hospital B		Hospital C		Hospital D		Hospital E		
		%	d	%	d	%	d	%	d	%	d	p (χ^2)
All patients	2700	**0.2**	519	**3.0**	740	**0.8**	518	**6.5**	540	**3.9**	383	<0.0001
By predicted risk of death at 6 months												
Lower tertile	902	**0.8**	130	**2.7**	224	**0**	146	**7.9**	279	**5.7**	123	<0.0001
Middle tertile	898	**0**	173	**3.8**	262	**1.1**	188	**7.7**	142	**3.8**	133	0.001
Upper tertile	900	**0**	216	**2.4**	254	**1.1**	184	**1.7**	119	**2.4**	127	0.228
Of patients treated with s.c. heparin, proportion with a CT head scan	77	**100**	1	**91**	22	**75**	4	**100**	35	**100**	15	Not calculated

Table 8.14 The process of stroke care: the Royal College of Physicians Stroke Audit Package (RCPSAP)

	Median proportion of RCPSAP criteria complied with % (IQR)										
	Hospital A n = 116		**Hospital B** n = 104		**Hospital C** n = 115		**Hospital D** n = 100		**Hospital E** n = 121		p (KW)
Overall	**52**	(44 to 65)	**62**	(56 to 70)	**65**	(57 to 79)	**79**	(67 to 88)	**67**	(57 to 79)	**<0.0001**
Per domain (no. of criteria)											
History (3)	100	(67 to 100)	100	(67 to 127)	100	(67 to 111)	100	(100 to 100)	100	(67 to 100)	0.001
Risk factors (8)	50	(25 to 75)	63	(38 to 75)	75	(38 to 88)	75	(50 to 100)	75	(38 to 100)	<0.0001
Pre-stroke function (3)	50	(0 to 100)	67	(0 to 100)	100	(46 to 100)	100	(67 to 100)	100	(50 to 100)	<0.0001
General examination (6)	67	(50 to 83)	67	(50 to 83)	83	(67 to 100)	83	(67 to 100)	67	(50 to 83)	<0.0001
Neuro. assessment (10)	67	(40 to 79)	67	(43 to 79)	71	(57 to 100)	85	(67 to 100)	70	(50 to 87)	<0.0001
Usual investigations (5)	80	(60 to 100)	100	(80 to 100)	100	(80 to 100)	100	(100 to 100)	100	(100 to 100)	< 0.0001
Special investigations (2)	100	(100 to 100)	100	(100 to 100)	100	(100 to 100)	100	(100 to 100)	100	(100 to 100)	0.049
Early management (4)	50	(25 to 63)	50	(50 to 75)	50	(50 to 75)	67	(50 to 75)	50	(50 to 75)	< 0.0001
Rehabilitation (6)	50	(17 to 67)	67	(50 to 83)	50	(33 to 67)	67	(67 to 83)	67	(50 to 83)	< 0.0001
Discharge planning (6)	25	(17 to 83)	33	(17 to 100)	67	(17 to 100)	83	(46 to 100)	50	(17 to 100)	< 0.0001
Secondary prevention (6)	67	(50 to 100)	100	(50 to 100)	100	(50 to 100)	100	(75 to 100)	100	(50 to 100)	0. 014
Ability to swallow (1) ‡	**17**†	(10 to 26)	**35**†	(25 to 45)	**45**†	(35 to 55)	**57**†	(46 to 67)	**47**†	(37 to 56)	<0.0001†
Clinical diagnosis (1) *	**11**†	(6 to 17)	**21**†	(13 to 29)	**24**†	(17 to 32)	**49**†	(39 to 59)	**22**†	(14 to 29)	<0.0001†

KW Kruskal Wallis test

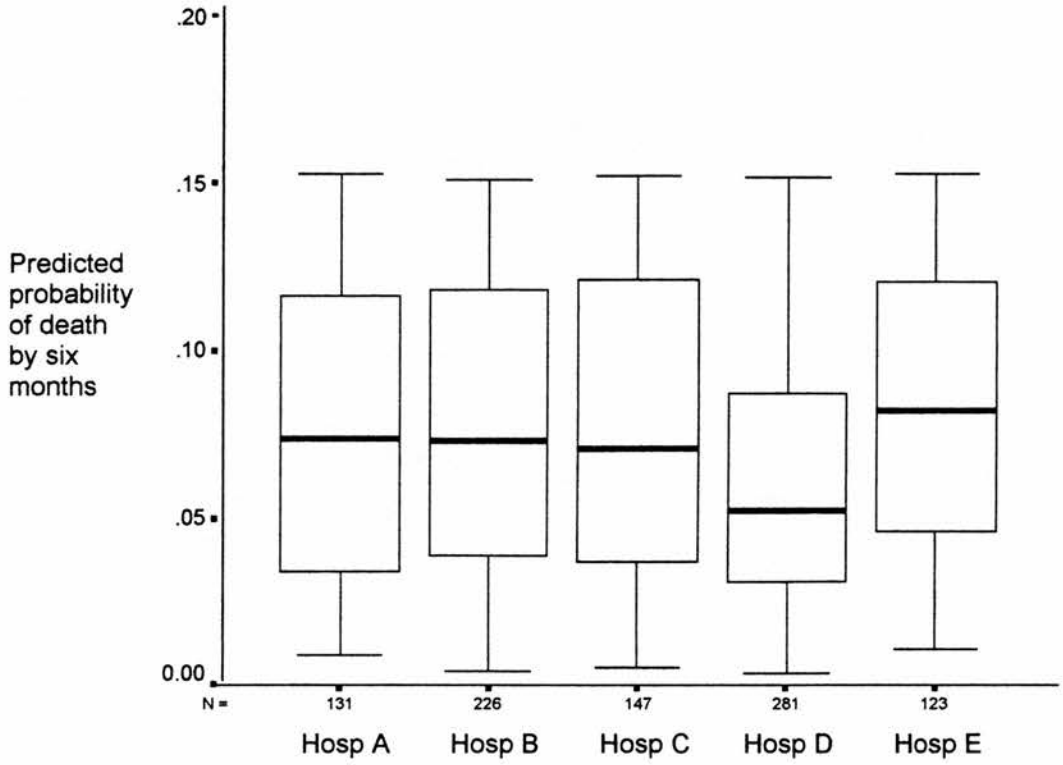
* Clinical diagnosis = a summary at the end of the clerking detailing the neurological deficit, likely site of cerebral lesion and relevant risk factors

† Only one criterion, therefore compliance reported as a proportion (with 95% confidence intervals); χ^2 test used to compare proportions

‡ Documented evidence that attention was paid to the ability to swallow is one of the audit criteria included in the neurological assessment domain

Figure 8.1 Comparison of the median predicted probability of death by six months (with inter-quartile range & 2.5th & 97.5th percentiles) between hospitals per tertile of predicted risk

1. Lower tertile of predicted risk



2. Middle tertile of predicted risk

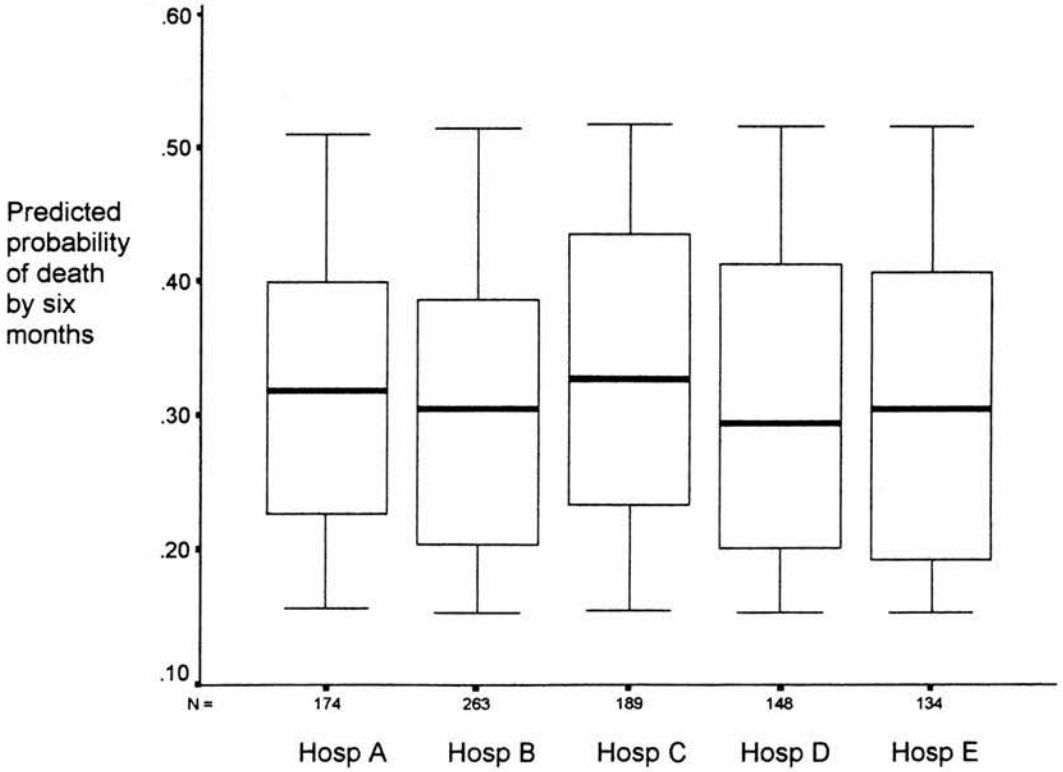
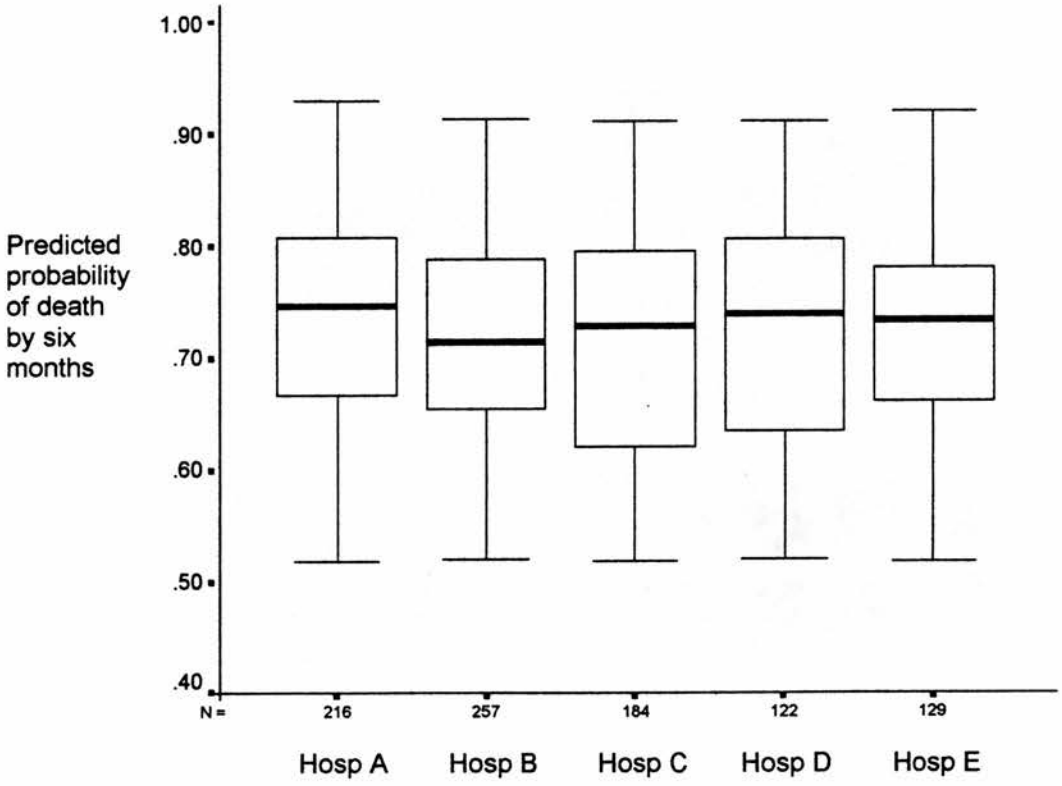


Figure 8.1 cont.

3. Upper tertile of predicted risk



Chapter Nine: Conclusion

In this thesis, I have attempted to address two questions regarding the use of outcomes data to measure the quality of stroke care. The first question was practical and asked whether it is feasible to considerably improve the systems currently used to collect outcome data for stroke in Scotland. The findings of this thesis suggest that this may indeed be the case, although a number of caveats still apply.

In Chapter Four, I showed that the strategy of selecting subsets of cerebrovascular disease codes in order to *routinely* identify cohorts of patients with acute stroke, whilst reasonably accurate at many hospitals, may result in seriously misleading comparisons of case fatality (in terms of attempts to measure the quality of care) at others. To improve the accuracy of routine case identification it is clear that the root cause of the problem, namely the poor quality of diagnostic information in discharge documentation, has to be addressed. This will require education of physicians, their involvement in coding and the provision of adequate time and resources. It might be argued that the strength of the commitment of the NHS to the collection of high quality clinical information might be measured by the extent to which these rather old suggestions are implemented. However, and perhaps fortunately, it is also clear that an important secondary benefit of a decision to collect our casemix variables is that they would effectively label the patient as having had an acute stroke. Hence, should they be collected, it is likely that the problem of biased case identification would diminish considerably.

In Chapter Five, I showed that our casemix variables were very reliable when measured prospectively, reasonably reliable and valid when measured retrospectively, just as accurate as urinary incontinence and conscious level, and that minor changes in their presentation might lead their collection to be more accurate still. Thus, it seems our casemix variables may indeed be suited to routine collection. However, I did not establish this directly (I did not measure the reliability and validity of the variables under truly routine conditions and in all grades of stroke) and ideally this should be done. Nonetheless, given their simplicity and the encouraging findings of the studies that were performed, it seems unlikely that their collection by junior doctors would be seriously in error.

In order to make the routine reporting of our casemix variables a reality, clinicians would first have to be persuaded to engage in their systematic collection. To do so, the task would have to be made simple and appear relevant (Kendrick 2001). The provision of a means of recording the variables as part of the usual process of care, for example on a stroke clerking proforma, would ensure that data collection was straightforward (Kendrick 2001). The relevance of the exercise would hopefully be obvious. Unlike the current system in Scotland which clinicians view as remote and of dubious value, our proposed system is grounded in rigorous clinical research and is capable of delivering credible comparisons of outcome. As such, it is likely that its very existence would tap into the natural enthusiasm of clinicians to participate in activities that allow them to better understand and improve their practice (Kendrick 2001). However, for clinicians not won over by this argument, it is also the case that data collection could be mandated as part of clinical governance.

In Chapters Four, Six and Seven, I showed that SMR1 data in their current format can be used to construct a system of postal follow up for survivors of stroke that would be practically feasible. Although feasible, it was also clear that the overall level of response would be poor after a single contact, only moderate after a second, and that the level of response would vary between hospitals. Despite this, and rather interestingly, bias in the comparison of functional outcome between several hospitals in the SOP remained small, especially if a second follow up was sent. However, between other hospitals, perhaps the minority, it is likely that bias would remain large enough to mislead efforts to measure the quality of stroke care. Fortunately, should the routine collection of functional outcome ever be attempted, it is likely that other strategies would reduce the impact of non-response to a more acceptable level: first, response might be improved by the perhaps obvious expedient of routinely reporting the *full* address to which the patient was *discharged* on the SMR1 and also by changes in the design of the contact letters and questionnaires; and second, adjustment for important casemix (using our methods) would itself attenuate the impact of non-response, although the extent of this 'protection' might vary between hospitals. This conclusion requires confirmation by further study.

In the SOP, the efficient operation of the system of follow up required only the services of a small number of part-time staff. One might argue, therefore, that a relatively small but dedicated unit might be able to do the same for the roughly 7,000 stroke survivors discharged from hospital every year in Scotland (Clinical Outcomes Working Group 1999b). Indeed, one might envisage an official centre having the ability to automatically link to other centrally held databases to identify up-dated

survival, GP and patient address data, which would greatly facilitate the entire process. Alternatively, modification of the SMR1, now an entirely electronic document, to report the address to which the patient was discharged should also be simple. In reality, however, the cost of establishing and running a routine system of postal follow up at a national level, coupled perhaps with its very novelty, seems likely to prevent its implementation in the near future.

The second question addressed asked whether our improved system of outcome measurement would be able to act as a valid indicator of the quality of stroke care. My findings in Chapters Seven and Eight raised the possibility that should our full system ever be implemented (one reporting case fatality and death or dependency) the resulting outcomes data might indeed have a role in identifying hospitals with moderately large differences in the quality of care and hence the system *might* have a role as a useful 'quality safety net'. On the other hand, should only our partial system be implemented (i.e. one which reported case fatality alone) it appears that it would have a more limited role, able to identify some hospitals with very marked shortcomings in care but would be insensitive to quite large differences in care between others. The additional reporting of institutionalisation after stroke (as a proxy for functional outcome) would not appear to rectify this situation. The explanation for this finding is likely to relate to the peculiar sensitivity of institutionalisation to baseline stroke severity and to social and cultural factors outside of hospital control, in addition, perhaps, to the crudeness of dichotomised residence data. Regardless, it is clear that differences between hospitals in rates of

institutionalisation should be interpreted with particular caution and their routine publication, at least as a quality indicator, perhaps reconsidered.

However, because of the small number of hospitals studied and the likely continuing influence of bias, confounding and the play chance, all the foregoing conclusions must remain tentative. Indeed, it is important to recall that virtually no other studies which have examined the relationship between outcome and the quality of stroke care using the hospital as the unit of analysis have reached similarly positive conclusions (see section 2.4). This observation alone raises the possibility that my findings may be spurious. On the other hand, it is also the case that the SOP was in many ways methodologically superior to previous studies: the SOP studied a larger number of hospitals and patients than the negative studies which also collected detailed casemix *and* process of care data; and the SOP made considerably more detailed measurements of casemix, outcome and the structure and process of care than the negative studies which included larger numbers of hospitals. On balance, therefore, it would seem prudent to regard the positive findings of the SOP simply as interesting possibilities that merit further investigation. To establish the true ability of our systems to indicate the quality of stroke care will require a much larger study involving tens of hospitals, able to determine their sensitivity and positive predictive value. As noted, it seems far from certain that this will ever be carried out.

In the absence of a definitive study, and assuming that my findings are true, it is still possible to make some predictions as to the likely accuracy and usefulness of our

improved systems (Davies and Crombie 1997; Mannion and Davies 2002; Palmer 1997). Unfortunately, it seems that their utility would be limited:

- Even if our full system were implemented, it is inevitable that it would misclassify a proportion of hospitals. If simulation studies are to be believed, despite making powerful adjustment for casemix, it is likely that the sensitivity and positive predictive value of the system would actually be quite low. As noted, even with perfect casemix adjustment and sample sizes as large as 900, sensitivity might reach only 68% and positive predictive value only 32% (see section 1.6.3) The very considerable problems associated with false positive and false negative identification of hospitals as 'quality outliers' have already been discussed (see section 1.6.4). These problems would be even greater should only our partial system be implemented.
- Because of the need to provide reasonably confident comparisons of outcome, the system would be able to report on the quality of care only every few years i.e. the relevance of its findings to current practice would be reduced. The use of a rolling period of observation (say, the previous three years) might allow the production of annual reports but clearly this would only be a partial solution.
- Although the system might identify the existence of a quality problem, it would not identify the remedy i.e. the system would be a prelude to further investigation rather than to action to improve the quality of care.
- No matter how accurate they eventually turn out to be, our improved systems would remain limited to the identification of "bad apples". For the great bulk of hospitals all the effort put in to the collection of data would fail to provide certain

information with which to act to improve the quality of stroke care. Indeed, as I showed in Chapter Eight, the systems would undoubtedly obscure moderate but important opportunities to do so at many (perhaps all) hospitals.

The last point underlines how even our improved systems would be out of step with emerging practice in quality improvement. The contemporary philosophy is moving away from the notion of dividing hospitals into those that pass or fail in relation to some quality standard and then applying remedial action in the group that fails (Laffel and Blumenthal 1989). Instead, borrowing from industrial quality science, it is now recognised that improvement is necessary at all hospitals regardless of their baseline level of performance; that the knowledge for this improvement comes from a detailed understanding of the complex systems involved in the provision of health care; and that in overall terms greater benefit results from incremental improvements in the quality of care at the majority of hospitals rather than revolution at the few i.e. by moving the mean of the bell curve of the performance distribution rather than by dealing with the minority in its tails (Berwick *et al.* 1992a; Davies and Lampel 1998; Laffel and Blumenthal 1989). Clearly, even perfectly accurate outcomes data cannot properly address this agenda.

Of course, a range of improvements could be tried in an attempt to make our proposed systems more accurate still. In particular, drawing from my findings and the current literature, the following methods should be explored:

- More refined measures of functional outcome might be used. In Chapter Six I showed that the MRS should probably be used in preference to the simple dependency question and suggested that the practicality of using even more sensitive measures of functional outcome and methods of comparison in the routine setting should be investigated.

- Adjustment for casemix might be improved by the addition of other (routinely collectable) variables. In particular, the addition of social deprivation to the model predicting death or dependency should be investigated further (Chapter Seven). The addition of urinary incontinence also merits further investigation, although my findings suggest that it is unlikely to be helpful (Chapter Seven).

- More fundamentally, instead of using a traditional method of regression, adjustment for casemix using a multilevel model (a method which recognises that elements within a system are often linked in a hierarchical fashion, e.g. patient, unit, hospital, society, each giving rise to variation that contributes to the whole) might allow the variation in outcome that is attributable to the hospital to be better teased out (Anonymous 2000).

- By allowing for a prior hypothesis about the performance of hospitals (perhaps based on their outcomes to date (Langford 1997)) a Bayesian approach to accounting for the play of chance might allow for more confident comparisons of outcome (and hence better discrimination between hospitals) than is possible using a frequentist approach (Bland and Altman 1998). The two methods certainly lead to quite different impressions of the quality of care, although which is the more valid remains to be determined (Austin and Naylor 2001).

- The use of control charts might allow for more easily interpreted comparisons of adjusted outcome (Mohammed *et al.* 2001). A control chart plots the mean number of adjusted outcomes against the number of patients treated (to form an upward sloping line) and then adds two further lines (control lines) usually three standard deviations above and below. The hospital outcomes are then marked and, for hospitals with outcomes outside of a control line a 'special cause' is suggested and further investigation indicated. The benefit of the method, therefore, would appear to lie in its ability to dispense with a morass of confidence intervals and to draw attention to variation that is highly unlikely (<0.3% likelihood) to be due to chance.

Nonetheless, it cannot be over-emphasised that no matter how well these improvements might increase the confidence with which differences in outcome can be attributed to differences in care, they cannot overcome the fundamental limitations of an outcomes-based system, namely, its failure to inform about quality of care at the majority of hospitals, its reduced relevance to current practice and its inability to identify necessary remedial action. Thus, there is good reason to believe that despite the effort and expenditure, the return in terms of valid identification of opportunities to improve care may not be worthwhile. Furthermore, it is notable that a system based on outcomes would also be relatively easy to subvert. For example, performance might be 'improved' by assigning stroke as a secondary diagnosis on the SMR1 for patients who die, ensuring that they do not appear in any audit; by assigning patients with TIA a stroke code, thereby improving the prognosis of the

measured cohort; and by systematically reporting casemix worse than is truly the case, improving outcome in relation to that predicted (Dennis 2000).

A system based upon the routine measurement of the structure and process of stroke care might answer many of the criticisms of an outcomes-based system. Almost by definition, it would be widely applicable. As I showed in Chapter Eight, each hospital in the SOP had shortcomings in one or more aspect of care that represented an opportunity to improve their stroke service. A system that measured structure and particularly process of care would have a number of other advantages (Crombie and Davies 1998; Mant 2001;Palmer 1997). In comparison to outcomes data, these data would be:

- More easily measured. Consider, for example, the difficulty in measuring functional outcome after stroke in comparison to the ease with which the medical record or SMR1 can identify whether and for how long a patient was admitted to a stroke rehabilitation unit.
- More easily interpreted. Causal ambiguity permeates the use of outcomes data. Process data, on the other hand, more directly reflect the care that was provided and are much less influenced by variation in casemix and hence the question of attribution is clear. Also, provided the worth of the process has been established, data can be interpreted in relation to the research evidence, avoiding the need to make comparisons between hospitals.
- More sensitive. Large numbers of observations are needed to overcome the play of chance in order to reveal differences in outcome that can be confidently

attributed to differences in care. Much smaller sample sizes are required for comparisons of the process of care. For example, it has been calculated that to confidently draw attention to a 31% difference in the use of effective interventions after myocardial infarction requires the measurement of outcome in 1350 patients at each hospital but the measurement of the process of care in only 27 (Mant and Hicks 1995). Process data therefore provide much more timely measurements of the quality of care.

- An indicator for action. Measurement of process often suggests the actions needed to improve care. For example, the observation that, say, only 60% of patients with an ischaemic stroke were discharged on aspirin would immediately direct attention to an investigation of why this were so.

It is also suggested that process data provide the only means to monitor and prevent 'near misses' (potentially serious deficiencies in care that uncommonly lead to serious adverse outcomes), to learn from otherwise overlooked outcomes (e.g. the death of a patient with an elderly patient with a severe stroke may still have been contributed to by poor quality of care) and to study unnecessary resource use (Crombie and Davies 1998).

These considerable benefits have not been lost on the stroke community in the UK who, since the early 1990s, have focused their efforts to measure quality of care on this approach (see section 2.3). As noted, the RCPSAP represented the first attempt to develop a comprehensive instrument to measure the quality of stroke care in the UK (Gompertz *et al.* 1994a). Its later incarnation, the Intercollegiate Stroke Audit

Package (ISAP), was developed in tandem with an authoritative set of clinical guidelines and incorporated sections devoted to the structure of care and the activities of therapists and nurses (Gompertz *et al.* 2001; Rudd *et al.* 2001b). Similarly sophisticated tools for measuring the structure and process of stroke care have recently been developed in the USA (Holloway *et al.* 2001; LaClair *et al.* 2001) and a more limited instrument has been tested in Holland (van Straten *et al.* 1997). The practicality and power of the approach has been shown by a series of national audits using the ISAP in England, Wales and Northern Ireland over the period 1998 to 2001 (Rudd *et al.* 1999; Rudd *et al.* 2001b; Rudd and Pearson 2002). Despite participation being voluntary and without administrative support, over 95% of acute trusts that admitted stroke patients took part in the most recent survey. Data were collected locally (from the medical records of 40 consecutive patients), collated and analysed centrally, and the results fed back within three months. The reports were confidential but allowed for anonymous comparison with other hospitals and were combined with regional meetings to discuss the findings and promote action. This relatively simple, timely, inclusive and non-threatening approach coincided with, and is likely to have been partly responsible for, considerable improvements in the quality of stroke care. A further national audit is planned for 2004 and it seems that the method may become the routine tool for quality assessment in England, Wales and Northern Ireland (Rudd and Pearson 2002).

However, there are important shortcomings to a purely structure and process based approach to measuring the quality of stroke care which were examined in Chapter Eight in relation to the quality of our own data. To recap, the difficulties essentially

stem from our still limited (although undoubtedly expanding) knowledge of what constitutes effective stroke care *and* the fact that most interventions that we know or suspect are effective are not recorded in a manner that describes what was done or with how much skill, and yet, for practical reasons, the measurement of the process of care must still be based upon routinely recorded data. As a result, much of what is measured refers to the completeness of record keeping in relation to non-therapeutic aspects of care and/or to aspects of care whose link with outcome may not be grounded in hard evidence; and, even where there is hard evidence of effectiveness (in particular in relation to the provision of formally organised stroke care) because of the difficulties of definition and measurement, one cannot be certain that the measured items are equally effective across different clinical settings and hence the meaning of comparisons is unclear.

To these fundamental problems can be added the fact that the collection of process data is time consuming – it is estimated that to audit 40 sets of medical records using the ISAP takes 20 hours (Rudd *et al.* 2001b) – and may be prone to observer error. Indeed, when self assessed, the potential for observer error might relatively easily translate into fraud, either through the simple expedient of relabelling current practice (such that an ordinary ward round becomes a MDT meeting and an ordinary ward becomes a stroke unit) or cheating in the completion of audit forms (Dennis 2000). Structure and process criteria are also prescriptive and so run the risk of stifling clinical innovation and, by reflecting current practice, have the potential to become obsolete (Mannion and Davies 2002). With our current knowledge base and methods, therefore, it would seem prudent to interpret differences in structure and

process of stroke care with some caution (Davenport and Dennis 1996). Indeed, given these facts and together with the finding that hospital level analyses have until now failed to associate better process scores with better outcomes, some observers have suggested that outcomes should remain the prime measure of quality until we better understand the process of care (McNaughton *et al.* 2003), which, of course, brings us back to where we started from.

Thus, stroke medicine finds itself in the difficult situation that neither a system based solely upon outcomes nor one based solely upon the structure and process of care can provide a completely valid and practical way to measure the quality of hospital stroke services. The practical solution to this dilemma would appear to lie in the observation that the advantages and drawbacks of each method are, in many respects, complimentary (Mannion and Davies 2002). The real challenge would therefore appear to lie in the creation of a system capable of measuring structure, process *and* outcome. A system such as this would almost certainly be more effective and acceptable. Consider, if both sets of measurement were to agree about the quality of care, it would be more difficult to dispute the findings and any suggested improvements would be more likely to be pushed through (Donabedian 1988); and, if the two methods were to disagree, attention would be directed to the possibility of inadequate or manipulated measurement (Donabedian 1988), helping to guard against simplistic or erroneous conclusions and making plain the complex reality of quality measurement.

Just such a system, known as "Riks-Stroke", has been in operation in Sweden since 1994 (Asplund *et al.* 2003). Data collection is kept simple to ease participation and includes clinically important baseline characteristics, key items of care and outcome at discharge and at three months (survival, functional status and residence). Hospitals receive confidential annual reports in which their results are compared with national data and suggestions made as to how to improve their service. The public are able to access the data but only at a county level i.e. aggregated across two or more hospitals. Participation is voluntary and the centre is in regular contact with hospitals, establishing and refining the areas of measurement. A drawback is that not all patients at each hospital are registered on Riks-Stroke, opening up the possibility of biased comparisons. Nonetheless, Riks-Stroke has coincided with and is likely to have contributed to clear improvements in stroke services in Sweden over the last ten years, particularly in the area of stroke unit provision. A comparable regionally based system has operated in Germany since 1994 (Heuschmann *et al.* 2003). More recently, national stroke registries capable of collecting structure, process and outcome data have been established in Austria (Steiner *et al.* 2003) and Poland (Czlonkowska *et al.* 2003) and, in the USA, a prototype system, the Paul Coverdell National Acute Stroke Registry (Wattigney *et al.* 2003), is under development.

The situation in Scotland has also moved on from that in the early 1990s when the SOP was first designed. Following the reorganisation of the NHS in 1997 and the report of the Coronary Heart Disease and Stroke Task Force in 2001 (Anonymous 2001), an integrated approach to tackling Scotland's heavy burden of cardio- and cerebrovascular disease is being developed (Anonymous 2002). The establishment

of a credible and effective national system to routinely measure and improve the quality of stroke care forms a central plank of this strategy. In line with this, the official emphasis on the value of outcomes data has been replaced with plans to create a system capable of routinely measuring the structure, process and outcome of stroke care. It is gratifying that the methods and findings of the SOP have made an important contribution to the development of this approach.

The foundation of the proposed system is a set of rigorously developed national standards and review criteria, based on the Cochrane database of systematic reviews and Scottish Intercollegiate Guidelines Network documents relating to the management of stroke (Anonymous 2003b). For each patient, hospitals will be required to collect a data set comprising simple aspects of the process of stroke care (including: whether the patient was admitted to a stroke unit, was under the care of a stroke physician, had a CT head scan or was discharged on an anti-thrombotic drug) and casemix data including the pathological sub-type of the stroke and the six variables used to adjust comparisons of outcome in the SOP. Appropriate computer software will be provided but the means of entering the data will remain the responsibility of each hospital. The data will be reported to the centre periodically, linked to centrally held death certification data, and comparative data made public. A study to test the feasibility of routinely collecting the proposed minimum data set is under way. In addition, selected aspects of the structure and process of care will be audited periodically in greater detail using a method similar to, but briefer than, the ISAP. Currently there are no plans to measure functional outcome and residence at a defined time after admission, although interestingly, the national strategy for stroke

mandates the follow up of all survivors at 3 months (Anonymous 2003b). This would appear to be an ideal opportunity to get around the whole issue of postal follow up and, instead, provide a questionnaire inquiring into functional outcome as part of the usual process of care, perhaps with arrangements for its return to a central authority by post to avoid the possibility of biased response.

Thus, having begun this thesis as a response to Scotland's brave but in some senses misguided attempt to routinely measure the quality of stroke care using a crude system of outcome measurement, I end it with Scotland on the verge of acquiring a sophisticated system capable of measuring all three elements of the structure, process and outcome triad. Indeed, by forcing the issue of quality of care to be debated and researched, this remarkable change may be the greatest legacy of the decision to publish the CRAG Outcome Indicators (Mannion and Davies 2002). Nonetheless, the development of a comprehensive and credible system of measurement is only part of the story. Long experience suggests that the use of measurement data to bring about real improvements in the quality of care depends also upon the environment into which those data are released. Their effective use requires both the creation of a learning culture where failure can be openly discussed and the establishment of systems within hospitals capable of responding in a constructive manner (Mannion and Davies 2002; Mannion and Goddard 2003). A key element in establishing a learning culture is the perception that the system for measuring performance is used for understanding rather than judging the activities of clinicians (Thomson *et al.* 1997). For the system proposed in Scotland, the close co-operation of clinicians in its development and their active involvement in data collection should help ensure a

feeling of ownership rather than oversight. However, as a centrally orchestrated and published system, care will have to be taken to prevent it from appearing to 'name and shame' under-performing units. Indeed, the question of routine publication of its findings may need to be revisited. Within hospitals, a number of inter-linked strategies will have to be developed to disseminate results to front line staff and to support the implementation of change. The latter might involve education of clinicians and other users, fostering networks of interested professionals to encourage the exchange of ideas, identification of clinicians to champion the data and to provide a clear direction for change, the provision of appropriate incentives (in particular those related to peer status and professional reputation), and perhaps even assistance in adopting some of the successful techniques of industrial quality science (Berwick *et al.* 1992a; Berwick *et al.* 1992b; Berwick 1996; Davies and Lampel 1998; Laffel and Blumenthal 1989; Mannion and Davies 2002; Mannion and Goddard 2003). The decision to create a series of managed care networks for stroke in Scotland and for those networks to talk to one another is a welcome step in this direction. However, in order to make best use of the information that is about to become available and to guard against any negative consequences of performance measurement, it is clear that much work remains to be done. Rather than further honing the tools of quality measurement, important though this may be, the most important challenges facing those wishing to actually *improve* the quality of care for patients with stroke in Scotland would now appear to lie in this direction.

Appendix 1: Medical record audit form

Discharging Hospital No

Date information collected :

Date sent to punching :

Date returned from punching :

HLT Stroke Data Collection Form

Surname of patient	
First Initial	
Address admitted from	
Post code	
Date of birth	/ /
Sex	*SMR1 code*
Marital state	*SMR1 code*
GP Name	
GP Address	
Post code	
Is data available for this patient ?	*YES = 1 NO = 2*
Date of first admission	/ /
Date of first clerking	/ /
Did this patient have an acute stroke ?	*YES = 1 NO = 2*

If NO what was the diagnosis ?

If diagnosis is not stroke. no more data need be collected

Discharging Hospital No

If **YES** to acute stroke - Date of symptom onset / /

Date unknown *YES = 1 NO = 2*

What category of stroke did patient have ? (Find from results of CT scan or autopsy)

Circle one of the following codes :

cerebral infarction	1
intracerebral haemorrhage	2
subarachnoid haemorrhage	3
not known	9

From which **hospital** was patient finally discharged ? *code*

Case reference number

Date of discharge / /

Type of discharge *SMR1 code*

Consultant (name)

Speciality *SMR1 code*

Address discharged to
(if different from admission address)

Post code

Discharging Hospital No	

From which **ward** was patient finally discharged ?

Type of ward

SMR1 code [ ]

Ward name/no .....

code [ ]

Hospital .....

Date of discharge [/ /]

*If patient was in another **ward** prior to this :*

Type of ward

SMR1 code [ ]

Ward name/no .....

code [ ]

Hospital .....

Date of discharge [/ /]

*If patient was in another **ward** prior to this :*

Type of ward

SMR1 code [ ]

Ward name/no .....

code [ ]

Hospital .....

Date of discharge [/ /]

*If patient was in another **ward** prior to this :*

Type of ward

SMR1 code [ ]

Ward name/no .....

code [ ]

Hospital .....

Date of discharge [/ /]

Discharging Hospital No

*If patient was in another **ward** prior to this :*

Ward name/no.....

Hospital.....

Type of ward

SMR1 code [ ]

code [ ]

Date of discharge [/ /]

*If patient was in another **ward** prior to this :*

Ward name/no.....

Hospital.....

Type of ward

SMR1 code [ ]

code [ ]

Date of discharge [/ /]

Discharging Hospital No.	

YES = 1. NO = 2

1. **Does the patient have a history of Diabetes Mellitus ?** ☐

2. **Does the patient have a history of :**

   (a) Ischaemic heart disease ? ☐

   (b) Myocardial infarction ? ☐

3. **Is the patient married, not a widow/er and not divorced/separated ?**
   or Does tne patient have a living partner ?
   or Does the patient NOT live alone ? ☐

4. **Is the patient known to be in employment ?**
   Where not known, record as NO ☐

5 Prior to the event, could the patient live independently i.e. carry out everyday activities unaided ? (e.g. walking, bathing, feeding, dressing) ☐

6 Is the patient's systolic blood pressure > 160 mmHg ? ☐

7 Does the patient open their eyes spontaneously WITHOUT verbal or painful stimulus ? ☐

8 Does the patient move their UNAFFECTED limbs (if any) purposefully ? This does not consider their affected limbs. ☐

9 Can the patient tell you their name, the place and time correctly ? ☐

10 Can the patient lift both arms against gravity ? ☐

11 Is the patient unable to walk without the aid of another person ?
They may use any other aid ☐

Discharging Hospital No.	

YES = 1, NO = 2

12. **Was a CT scan done ?** ☐

If yes, **date of first scan.** / /

13. **Was the patient given subcutaneous heparin ?** ☐

If yes, **date started.** / /

14. **Was the patient given parenteral fluids ?** ☐

If yes, **date parenteral fluids started.** / /

15 Was an NG tube inserted ? ☐

If yes, date first inserted. / /

Was this for (circle one code)
- fluids only ? 1
- fluid & food ? 2
- neither/uncertain ? 9

16 Has urinary incontinence been recorded since the event ? ☐
(Patient has an indwelling catheter or 2 or more documented episodes of incontinence within 7 days of the event)

17 Was a urinary catheter inserted ? ☐

If yes, date inserted / /

18 Was a multi-disciplinary case conference held ? ☐

If yes, date of first multi-disciplinary case conference / /

19 Was a PEG tube inserted ? ☐

If yes, date first inserted / /

Discharging Hospital No	

YES = 1, NO = 2

20. Did the patient have any fractures

If yes, site of most recent fracture..... *LSR code*

date most recent fracture was noted / /

site of next most recent fracture..... *LSR code*

date fracture noted / /

site of next most recent fracture..... *LSR code*

date fracture noted / /

21. Did the patient have any pressure sores ?

If yes, date pressure sore first noted / /

22 Was the patient finally discharged on Aspirin ?

23 Was the patient finally discharged on Warfarin ?

Is there any other relevant information about patient history ?

Please state :

Details of time taken by auditor : mins

Were any problems encountered by auditor ?

(Detail overleaf)

RCP ?

Appendix 2: Royal College of Physicians Stroke Audit Package

CONFIDENTIAL

Royal College of Physicians Stroke Audit Form

Surname | First Name | Code Number

Please delete or white out the patients name after the audit has been completed.

1: History

24hrs

	Yes	No	No, but...
Is the source of the history documented?	☐	☐	☐

Answer "no, but..." if it is recorded that patient is alert & oriented and communication is normal.

Is the rate of onset of symptoms recorded?	☐	☐	☐
Any record of drugs on admission? *(dose & freq unless PRN)*	☐	☐	☐

Answer "no, but..." if it is recorded that patient is unconscious, confused, aphasic or unaware of the symptoms and there are no possible witnesses.

2: Risk factors

7 days

Is the presence or absence of the following risk factors documented?
("nil relevant", "previously well", "PMH nil" are not acceptable)

	Yes	No	No, but...
Previous stroke or transient ischaemic attack	☐	☐	☐
History of hypertension	☐	☐	☐
History of heart disease *(angina, MI, valves, arrhythmias, failure)*	☐	☐	☐
Peripheral vascular disease *(claudication, arterial surgery)*	☐	☐	☐
History of diabetes	☐	☐	☐
History of hyperlipidaemia	☐	☐	☐
Smoking *(non, ex, or current)*	☐	☐	☐
Alcohol *(must specify amount in units, pints, bottles etc.)*	☐	☐	☐

Answer "no, but..." if it is recorded that patient is unconscious, confused or aphasic and died within 3 days. Where survivors cannot communicate, details should have been obtained from another source, e.g. GP within 7 days.

CONFIDENTIAL

72hrs

3: Pre-stroke function

	Yes	No	No, but...
Is ONE of A, B or C recorded?	☐	☐	☐

A. Two or more activities of daily living

(e.g. Ambulation, Bathing, Continence, Dressing, Emotion/cognition)

B. Any more difficult tasks *(e.g. shopping, driving, working)*

C. That the patient is either "housebound" or "independent" *("Fit and well" is not adequate)*

☞ *Answer "no, but..." if it is noted that information not available on admission and patient died within 3 days.*

Is use of social services recorded?	☐	☐	☐

☞ *Answer "no, but..." if it is noted that information not available on admission and patient died within 3 days, or noted to be independent or performing more difficult tasks.*

Is employment recorded?	☐	☐	☐

☞ *Answer "no, but..." if it is noted that information not available on admission and patient died within 3 days, or is over 65 years old.*

24hrs

4: General examination

Have the following been recorded?	Yes	No
Pulse rate and rhythm	☐	☐
Blood pressure	☐	☐
Heart sounds	☐	☐
Neck bruits *(presence or absence)*	☐	☐
Peripheral pulses *(presence or absence)*	☐	☐
Fundoscopy *(at least that it was attempted)*	☐	☐

CONFIDENTIAL

24hrs

5: Neurological assessment

Have the following been specifically recorded?	Yes	No	No, but...
Conscious level *(e.g. Glasgow Coma Scale (GCS), alert, drowsy, unrousable)*	☐	☐	
Eye movements *(III IV VI intact, doll's eye response if unconscious)*	☐	☐	
Power in the limbs *(response to pain if unconscious)*	☐	☐	
Communication *(recorded as normal or any abnormality specified)*	☐	☐	☐
Trunk control or gait	☐	☐	☐
Swallowing *(gag reflex is not enough)*	☐	☐	☐

☞ *Answer "no, but..." if it is recorded that patient is unconscious.*

	Yes	No	No, but...
Formal mental test score *(e.g. Hodkinson)*	☐	☐	☐
Visuospatial function *(e.g. neglect, inattention, agnosia etc.)*	☐	☐	☐
Visual fields	☐	☐	☐
Sensory testing *(including proprioception)*	☐	☐	☐

☞ *Answer "no, but..." if it is recorded that patient is unconscious/drowsy (GCS < 15), unable to communicate, or otherwise unable.*

24hrs

6: Clinical diagnosis

	Yes	No
Has a clear diagnostic formulation been made? *(including neurological deficit, likely site of cerebral lesion and relevant risk factors)*	☐	☐

7 days

7: Usual baseline investigations

Are these in the notes?

☞ *Answer "no, but..." if patient died within 24 hours, a decision to withhold active treatment noted, or investigation performed within previous month.*

	Yes	No	No, but...
FBC	☐	☐	☐
ESR/Viscosity	☐	☐	☐
U&E	☐	☐	☐
Glucose	☐	☐	☐
ECG	☐	☐	☐

CONFIDENTIAL

7 days

8: Special investigations

Have any of the following been carried out?

	Yes	No	No, but...
CT Scan	☐	☐	☐

☞ *Answer "no, but..." if it is recorded that:-*
a) an accurate diagnosis is not needed to make management decisions
Or,
b) none of the following "cast-iron" indications for CT apply
i) diagnosis is in doubt - *no clear onset or gradual onset over 48hrs+*
- *unexpected deterioration after stroke*
- *unexplained unconsciousness*
ii) cerebellar symptoms or signs
iii) marked headache, vomiting or meningism
iv) patient on warfarin, or anticoagulation is considered
v) carotid endarterectomy is considered

	Yes	No	No, but...
Investigation for rare causes of stroke *(e.g. lupus anticoagulant, VDRL, ANF, echo, blood cultures, clotting studies)*	☐	☐	☐

☞ *Answer "no, but..." if patient is over 55 or one or more risk factors are documented*

24hrs

9: Immediate management plan

Are plans for the following recorded?

	Yes	No	No, but...
Hydration *(records must specify oral, IV + Nil By Mouth, or nasogastric)*	☐	☐	☐

☞ *Answer "no, but..." if swallowing is said to be normal*

Urinary incontinence *(records must specify pads, sheath or catheter)*	☐	☐	☐

☞ *Answer "no, but..." if it is recorded that patient is continent of urine*

7 days

10: Management during the first week

Any record of the following?

	Yes	No	No, but...
Consultant review	☐	☐	☐
Review of important neurological deficits *(i.e. conscious level, communication, swallowing, hemiparesis)*	☐	☐	☐

☞ *Answer "no, but..." if patient dies within 24 hours*

CONFIDENTIAL

1 month

11: Rehabilitation

Any record of the following in the first month? *(in the medical notes)*	Yes	No	No, but...
Personal interests	☐	☐	☐
List of patient's problems *(functional & social if appropriate)*	☐	☐	☐
Objectives, or goals of rehabilitation *(destination at least)*	☐	☐	☐
Re-assessment of functional status	☐	☐	☐
The results of a multidisciplinary meeting	☐	☐	☐

☞ *Answer "no, but..." if patient dies or is discharged within 7 days*

Information given to patient and/or relatives	☐	☐	☐

☞ *Answer "no, but..." if patient unconscious, receptive dysphasia or confused and there are no relatives.*

During admission

12: Discharge planning

Are the following recorded? *(In notes or discharge summary)*	Yes	No	No, but...
Ownership of accommodation *(Council, owner occupied)*	☐	☐	☐
Type of accommodation *(e.g. residential, warden-supervised, private)*	☐	☐	☐
Whether living alone or not	☐	☐	☐
Stairs/ground floor/lift	☐	☐	☐
Access to toilet	☐	☐	☐
Whether informal support is available *(e.g. friends, relatives, neighbours)*	☐	☐	☐

☞ *Answer "no, but..." if any of the following apply; it is recorded that the information was not available, and the patient died or was discharged within 7 days; the patient was in residential accommodation; or a full functional recovery is documented.*

CONFIDENTIAL

By 1st follow-up

13: Secondary prevention

Any record of the following?
(in the doctors' notes or summary, not cardex/nursing charts)

	Yes	No	No, but...
Blood pressure at least four days after stroke	☐	☐	☐
Anti-hypertensive medication	☐	☐	☐

☞ *Answer "no, but..." if patient dies or is normotensive (systolic <180 and diastolic <100)*

Long-term aspirin	☐	☐	☐

☞ *Answer "no, but..." if patient dies, haemorrhage is recorded or pt unable to take aspirin for any other reason.*

Long-term anticoagulation	☐	☐	☐

☞ *Answer "no, but..." if patient dies, is documented not to be in atrial fibrillation or any contraindication, such as haemorrhage, falls or confusion is recorded.*

Advice about smoking	☐	☐	☐

☞ *Answer "no, but..." if patient dies or is recorded to be a non-smoker.*

Non-invasive carotid imaging *(eg. duplex scan)*	☐	☐	☐

☞ *Answer "no, but..." if residual disability, haemorrhagic stroke posterior circulation stroke, stroke of uncertain distribution, patient unwilling to consider surgery, dies or aged >75.*

Appendix 3: Modified Rankin Scale

0	No symptoms
1	Minor symptoms which do not interfere with lifestyle
2	Symptoms which lead to some restriction of lifestyle but do not interfere with the patient's capacity to look after themselves
3	Symptoms which significantly restrict lifestyle and/or prevent totally independent existence
4	Symptoms which clearly prevent independent existence though not needing constant attention
5	Severe handicap, totally dependent, requiring constant attention day and night

Appendix 4: Outcome questionnaire

STRICTLY CONFIDENTIAL
M125M1EH667BE 19 April 1996

Please answer the following questions.

If you are unable to fill in this form yourself, please ask someone who knows you well to help you.

Please tick ***EITHER*** Yes ***OR*** No for every question.

1. Were you admitted to hospital with

a.	A heart attack?	Yes	☐	No	☐
b.	A stroke (including sub-arachnoid haemorrhage)?	Yes	☐	No	☐
c.	Another problem?	Yes	☐	No	☐

2. Where are you staying now?

a.	Same place as before your illness.	Yes	☐	No	☐
b.	In hospital.	Yes	☐	No	☐
c.	With family.	Yes	☐	No	☐
d.	In a nursing or residential home.	Yes	☐	No	☐
e.	Other, please specify	Yes	☐	No	☐

.....

3. Are you taking regular

a.	Aspirin?	Yes	☐	No	☐
b.	Blood pressure lowering tablets?	Yes	☐	No	☐
c.	Warfarin?	Yes	☐	No	☐
d.	Tablets for epilepsy, fits or seizures?	Yes	☐	No	☐

STRICTLY CONFIDENTIAL
M125M1EH667BE 19 April 1996

4. Do you need help from anybody with everyday activities? Yes ☐ No ☐

If you had a stroke

5. If you had a stroke (including sub-arachnoid haemorrhage), has it left you with any problems? Yes ☐ No ☐

Now, for each of the following five areas, could you please tick the one sentence which best describes how you are now.

a. Walking

I have no problems with walking about. ☐
I have some problems with walking about. ☐
I am confined to bed. ☐

b. Self care

I have no problems with self care. ☐
I have some problems with self care. ☐
I am unable to wash and dress myself. ☐

c. Usual activities (for example - work, housework, family or leisure activities).

I have no problems performing my usual activities. ☐
I have some problems performing my usual activities. ☐
I am unable to perform my usual activities. ☐

STRICTLY CONFIDENTIAL
M125M1EH667BE 19 April 1996

d. Pain/discomfort

I have no pain or discomfort.	☐
I have moderate pain or discomfort.	☐
I have extreme pain or discomfort.	☐

e. Anxiety /depression

I am not anxious or depressed.	☐
I am moderately anxious or depressed.	☐
I am extremely anxious or depressed.	☐

Finally, who completed this questionnaire?

The patient	☐
Friend/relative/carer	☐

Signature ..... Date .....
(of person completing the form)

Please return the completed questionnaire to us in the enclosed freepost envelope.

THANK YOU FOR YOUR HELP

Appendix 5: Follow up pack sent to General Practitioner

Date

«GPNAM»
The Surgery
12 Fishers Wynd
Brigadoon
EH66 IYU

Dear Dr Smith

Re: Mrs M Jones DOB: 12 December 1935

We understand from Dr McDowall that Mrs Jones was discharged from the Western General Hospital on the 19 April 1996 having had a stroke.

The physicians at the Western General Hospital are collaborating with us to determine the feasibility of following up all admitted stroke patients to provide information which might reflect the quality of hospital based stroke services.

We would be grateful if you or a member of your staff could check that this patient is still alive and forward the enclosed stamped envelope to them at their home address. Could you please complete the address on this envelope as we do not routinely have access to this information.

The envelope contains an introductory letter, a short questionnaire to determine the patient's functional status and a freepost envelope in which the patient or carer can return the completed questionnaire to us.

If the patient has died or is no longer on your list, please fill in any available details on the form attached and return it to us in the other freepost envelope enclosed. If, however, you are able to forward the enclosed envelope to the patient, please ignore the attached form.

The information which we are collecting will be kept strictly confidential.

Thank you very much for your help.

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant in Stroke Medicine

Please do not send this to the patient

If the patient has died or is no longer on your list, please fill in any available details on this form and return it to us in the freepost envelope which is enclosed with the patient's questionnaire.

If you have forwarded the questionnaire to the patient, please ignore this form.

Re: Mrs M Jones DOB: 12 December 1935

Has the patient died (please circle) ? ***Yes No Not known***

If yes When? .....

Where? .....

Cause of death .....

If the patient has moved practice, could you please tell us the name and address of the new GP.

.....

.....

.....

THANK YOU FOR YOUR HELP

Appendix 6: Follow up letter sent to patient

Date

Dear Mrs Jones

We understand from Dr McDowall that you were discharged from the Western General Hospital in April 1996.

We are studying how to use information about recovery to improve hospital care for stroke patients in the future. I am writing to ask if you would be kind enough to help in this study by letting us know about your recovery. Your reply is important and by contributing to this study, it may help to improve stroke care.

Could we ask you, or someone who knows you well, to complete and return the enclosed questionnaire? No stamp is required on the envelope provided and, of course, all information received will be kept strictly confidential. Many thanks for your help.

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant Physician

Appendix 7: Multiple logistic regression

The standard multiple logistic regression equation is given as:

$$ln\,[p / 1\text{-}p] = a + b_1X_1 + b_2X_2 + b_3X_3 + b_iX_i$$

where

ln is the natural logarithm

p is the probability of an outcome

a is a constant

X_i are the predictor variables and

bi are the coefficients of the predictor variables.

It follows that:

$$p / 1\text{-}p = e^Y$$

where

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_iX_I \text{ (= the linear predictor)}$$

and hence that:

$$p = e^Y / (1 + e^Y)$$

Appendix 8: W and Ws scores: calculation of standard error and confidence interval

W score

The standard error of the W score is calculated as:

$$se(w) = (100/n)\ \sqrt{\Sigma p_i\,(1-p_i)}$$

where

p_i is the probability of an outcome for the i^{th} individual

and hence the 95% confidence interval is calculated as:

$$W\ score +/- 1.96\ se$$

Ws score

The variance of w_j (the W score in each stratum) is calculated as:

$$var\ (w_j) = \frac{(\Sigma [p_i\,(1-p_i)])_j}{(n_j / 100)^2}$$

where

$(\Sigma [p_i\,(1-p_i)])_j$ is the sum of $p_i\,(1-p_i)$ in stratum j

The standard error of the Ws score is calculated as:

$$se\ (w_j) = \sqrt{\Sigma_{over\ strata}\,(var\ w_j)\,f_j^{\,2}}$$

where

$f_j^{\,2}$ is the square of the fraction of the reference population in interval j

and hence the 95% confidence interval is calculated as:

$$Ws\ score +/- 1.96\ se$$

Appendix 9: Method for estimating the error in the measurement of outcome that results from reduced sensitivity in the identification of acute strokes (using routine hospital discharge data)

Assumptions

1. Let the true case fatality rate be 30%
2. Let the true predicted case fatality rate be 30%, hence the ratio of observed to predicted deaths (O/P) = true adjusted case fatality = 1
3. Let the sensitivity of the routine hospital discharge data be: 90, 80 or 70%
4. Let the unreported ('missed') cohort of patients have varying

- baseline stroke severity (by predicted risk of death)

  examine scenarios where the mean baseline severity in the unreported cohort is the same (30% predicted risk of death) or worse (40 and 50%) than in the reported cohort.

- quality of care (by impact on O/P ratio of deaths; worse care gives higher ratio)

  examine scenarios where the quality of care in the unreported cohort is the same (O/P = 1) or worse (resulting in a 10% or 20% higher case fatality rate; O/P 1.10 and 1.20) than in the reported cohort.

Calculations

1. Take a hospital sample with 1000 patients.
2. Calculate the overall observed number of deaths: 1000 x 0.3 = 300
3. Calculate the overall predicted number of deaths: 1000 x 0.3 = 300
4. Calculate the number of patients in the reported and unreported cohort e.g. for 90% sensitivity = 900 and 100 patients, respectively.
5. Calculate the predicted number of deaths in the unreported cohort by applying the specified level of baseline severity e.g. for 40% predicted risk of death: 100 x 0.4 = 40.
6. Calculate the observed number of deaths in the unreported cohort by applying the specified O/P death ratio to the predicted number of deaths e.g. for 10% higher case fatality 40 x 1.10 = 44.

7. Calculate the predicted number of deaths in the reported cohort by subtracting the number of deaths predicted in the unreported cohort from the overall number predicted e.g. 300 – 44 = 256 deaths.

8. Calculate the observed number of deaths in the reported cohort by subtracting the number of deaths observed in the unreported fraction from the overall number of deaths observed e.g. 300 – 40 = 260.

9. Use the resulting data to calculate the crude and adjusted (O/P) case fatality rate in the reported cohort. Continuing the example:

   crude case fatality = 260/900 = 28.9%

   adjusted case fatality (O/P) = 260/256 = 1.02

Appendix 10a: Follow up letters sent to the GP and patient in the DIRECT arm of the follow up trial

i. Letter sent to GP to ensure passive consent to contact patient
ii. First letter sent to all patients
iii. Second letter sent to non-responders to first follow up

Date

«GPNAM»
The Surgery
12 Fishers Wynd
Brigadoon
EH66 IYU

Dear Dr Smith

Re: Mrs M Jones DOB: 12 December 1935

We understand from Dr McDowall that Mrs Jones was discharged from the Western General Hospital on the 19 April 1996 having had a stroke.

The physicians at the Western General Hospital are collaborating with us to determine the feasibility of following up all admitted stroke patients to provide information which might reflect the quality of hospital based stroke services.

We intend to send a short questionnaire, designed to determine functional status, directly to Mrs Jones using the address in her hospital record. To avoid distressing bereaved relatives, we will only mail questionnaires to patients who do not appear on monthly mortality records for your area. The information we are collecting will be kept strictly confidential.

If there is any reason why we should not send Mrs Jones a questionnaire or you wish to know more about the study, please phone us (and ask for Dr Nicolas Weir or Mr Mike McDowall) or return the enclosed form (freepost) within one week of the postmark on this letter. Please could you let us know the reason if we should not make contact – this information should prove helpful in future planning.

Thank you for your help.

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant in Stroke Medicine

Follow Up Refusal Form

Please do not contact:

Name:

Address:

The reason is:
(if dead, please state date, place and cause if possible)

Name and signature of General Practitioner

Date

Mrs M Jones
23 The High Street
Brigadoon
Edinburgh
EH66 4UU

Dear Mrs Jones

We understand from Dr McDowall that you were discharged from the Western General Hospital in April 1996.

We are studying how to use information about recovery to improve hospital care for stroke patients in the future. I am writing to ask if you would be kind enough to help in this study by letting us know about your recovery. Your reply is important and by contributing to this study, it may help to improve stroke care.

Could we ask you, or someone who knows you well, to complete and return the enclosed questionnaire? No stamp is required on the envelope provided and, of course, all information received will be kept strictly confidential. Many thanks for your help.

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant Physician

Date

Mrs M Jones
23 The High Street
Brigadoon
Edinburgh
EH66 4UU

Dear Mrs Jones

About a month ago, I wrote to ask how you had recovered following your discharge from the Western General Hospital in April 1996. As we've not heard from you so far, I thought I would write again.

We are studying how to use information about recovery to improve hospital care for stroke patients in the future. We are very keen to learn about your recovery and I hope you can help us in our study. Your reply will certainly make our study more accurate and increase the chance that our findings will influence future stroke care.

Could we ask you, or someone who knows you well, to complete and return the enclosed questionnaire? No stamp is required on the envelope provided and, of course, all information received will be kept strictly confidential. Many thanks for your help.

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant Physician

Appendix 10b: Follow up letters sent to GP and patient in the INDIRECT arm of the follow up trial

i. Letter sent to GP to request that they forward our questionnaire to the patient
ii. First letter to patient
iii. Second letter sent to GP to request that they forward our questionnaire to non-responders to first follow up
iv. Second letter to non-responding patients

Date

«GPNAM»
The Surgery
12 Fishers Wynd
Brigadoon
EH66 IYU

Dear Dr Smith

Re: Mrs M Jones DOB: 12 December 1935

We understand from Dr McDowall that Mrs Jones was discharged from the Western General Hospital on the 19 April 1996 having had a stroke.

The physicians at the Western General Hospital are collaborating with us to determine the feasibility of following up all admitted stroke patients to provide information which might reflect the quality of hospital based stroke services.

We would be grateful if you or a member of your staff could check that this patient is still alive and forward the enclosed stamped envelope to them at their home address. Could you please complete the address on this envelope as we do not routinely have access to this information.

The envelope contains an introductory letter, a short questionnaire to determine the patient's functional status and a freepost envelope in which the patient or carer can return the completed questionnaire to us.

If the patient has died or is no longer on your list, please fill in any available details on the form attached and return it to us in the other freepost envelope enclosed. If, however, you are able to forward the enclosed envelope to the patient, please ignore the attached form.

The information which we are collecting will be kept strictly confidential.

Thank you very much for your help.

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant in Stroke Medicine

Please do not send this to the patient

If the patient has died or is no longer on your list, please fill in any available details on this form and return it to us in the freepost envelope which is enclosed with the patient's questionnaire.

If you have forwarded the questionnaire to the patient, please ignore this form.

Re: Mrs M Jones DOB: 12 December 1935

Has the patient died (please circle) ? Yes No Not known

If yes When? .....

Where? .....

Cause of death .....

If the patient has moved practice, could you please tell us the name and address of the new GP.

.....

.....

.....

THANK YOU FOR YOUR HELP

Date

Dear Mrs Jones

We understand from Dr McDowall that you were discharged from the Western General Hospital in April 1996.

We are studying how to use information about recovery to improve hospital care for stroke patients in the future. I am writing to ask if you would be kind enough to help in this study by letting us know about your recovery. Your reply is important and by contributing to this study, it may help to improve stroke care.

Could we ask you, or someone who knows you well, to complete and return the enclosed questionnaire? No stamp is required on the envelope provided and, of course, all information received will be kept strictly confidential. Many thanks for your help.

Yours sincerely,

M.S. Dennis

Dr Martin Dennis
Consultant Physician

Date

«GPNAM»
The Surgery
12 Fishers Wynd
Brigadoon
EH66 IYU

Dear Dr Smith

Re: Mrs M Jones DOB: 12 December 1935

A month ago I wrote to asking if Mrs Jones was still alive following a stroke in April 1996. We have not had a reply from Mrs Jones and I hope you don't mind us chasing things up.

In collaboration with the physicians at the Western General Hospital we are studying the feasibility of following up stroke patients to provide information which might reflect the quality of hospital stroke services. The questionnaire has been sent to a relatively small number of patients so it is crucial that we achieve a high response rate to prevent potentially biased and misleading results.

In case you did not receive the previous letter or it got mislaid, I am enclosing a further copy. Even if you did pass on the previous letter, please could you do so again as similar problems may have prevented Mrs Jones replying. The enclosed envelope contains an introductory letter, a short questionnaire and a freepost envelope in which the patient or carer can return the completed questionnaire.

We would be grateful if you or a member of your staff could check that this patient is still alive and forward the enclosed stamped envelope to them at their home address. Could you please complete the address on this envelope as we do not routinely have access to this information. If the patient has died or is no longer on your list, please fill in any available details on the form attached and return it to us in the freepost envelope enclosed.

The information we are collecting will be kept strictly confidential.

Thank you again for your co-operation

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant in Stroke Medicine

Please do not send this to the patient

If the patient has died or is no longer on your list, please fill in any available details on this form and return it to us in the freepost envelope which is enclosed with the patient's questionnaire.

If you have forwarded the questionnaire to the patient, please ignore this form.

Re: Mrs M Jones DOB: 12 December 1935

Has the patient died (please circle) ? Yes No Not known

If yes When? .....

Where? .....

Cause of death .....

If the patient has moved practice, could you please tell us the name and address of the new GP.

.....

.....

.....

THANK YOU FOR YOUR HELP

Date

Dear Mrs Jones

About a month ago, I wrote to ask how you had recovered following your discharge from the Western General Hospital in April 1996. As we've not heard from you so far, I thought I would write again.

We are studying how to use information about recovery to improve hospital care for stroke patients in the future. We are very keen to learn about your recovery and I hope you can help us in our study. Your reply will certainly make our study more accurate and increase the chance that our findings will influence future stroke care.

Could we ask you, or someone who knows you well, to complete and return the enclosed questionnaire? No stamp is required on the envelope provided and, of course, all information received will be kept strictly confidential. Many thanks for your help.

Yours sincerely,

M S Dennis

Dr Martin Dennis
Consultant Physician

Appendix 11: Ethical approval for the follow up trial

LOTHIAN RESEARCH ETHICS COMMITTEE

CERTIFICATE OF ETHICAL OPINION

LREC Reference Number: LREC/1997/4/52
Title: Improving response rates and estimating the non-response bias of a postal questionnaire used to follow up stroke patients admitted to hospital six months previously and discharged alive
Researcher: Dr N Weir

The Medicine/Clinical Oncology Research Ethics Committee of the Lothian Research Ethics Committee (the Committee) reviewed this proposed research and is of the opinion that it is ethical and appropriate to be carried out in the Lothian Area. This opinion encompasses all aspects of the application including the Patient/Subject Information Sheet and all other accompanying documentation provided.

The LREC application form, protocol, subject information sheet, information on compensation arrangements, payments to researchers and the provision of expenses to subjects (where appropriate) were reviewed and approved and the members of the Committee present at the meeting are shown on the attached *Membership List*.

This opinion is issued subject to the following conditions and is invalid if they are not followed:

- You must obtain appropriate management approval from the relevant NHS Trust(s) before starting the proposed research. It is the NHS Trust(s) that ultimately decide whether or not this research should go ahead taking account of the advice of the Local Research Ethics Committee.
- You must notify the Sub-Committee and the relevant NHS Trust(s), in advance, of any significant proposed deviation from the original protocol or application form and obtain approval for any such amendments using the *Amendment Approval Request Form.*
- You must submit reports to the Sub-Committee and the NHS Trust(s) once the study is underway if there are any unusual or unexpected results which raise questions about the safety of the research.
- You must report annually on successes, or difficulties, in recruiting subjects in order to provide useful feedback on perceptions of the study among patients and volunteers using the *Progress Report Form.*
- Where the study is terminated prematurely you must report within fifteen days indicating the reasons for early termination.
- You must submit a final report within three months of the completion of the study using the *Progress Report Form.*
- This opinion does not cover the inclusions of adults with incapacity in any study. Such opinion can only be given by the Multi-Centre Research Ethics Committee for Scotland.

Peter Reith
Secretary
Lothian Research Ethics Committee

Dale Keir
Administrator
Medicine/Clinical Oncology I
Research Ethics Committee

19 December 2003

Appendix 12: Statement of contribution

The collection of data for the SOP was carried out between August 1995 and March 1999 under the auspices of the Department of Clinical Neurosciences in the University of Edinburgh. I was a clinical research fellow employed in this department between September 1996 and March 2000 under the supervision of Dr. Martin Dennis; subsequently I have been employed as a Specialist Registrar in Neurology at the Royal Hallamshire Hospital, Sheffield. I performed all analyses between summer 1999 and winter 2002.

My contribution to the thesis was as follows:

I designed the sub-studies pertaining to:

- the accuracy of routine cerebrovascular disease discharge data;
- the validity and reliability of the collection of predictive data
- the methods of delivering the outcome questionnaire
- the ascertainment of outcome in non-responders to follow up
- the second survey of the structure of stroke care

I collected predictive data prospectively in 92 patients and retrospectively (from the medical records) in 200 patients with acute stroke.

I obtained ethical approval for the study investigating the response to different methods of postal follow up, supervised its running, and traced and obtained outcome data for the 66 non-responders.

I was responsible for dealing with day to day problems related to the running of the SOP proper.

I lead the task of cleaning, organising and checking the SOP data-set once data collection was complete.

I planned and carried out all the analyses presented, including the application of the appropriate statistics.

Appendix 13: Publications arising from this thesis

Publications

The reliability of the variables in a new set of models that predict outcome after stroke
NU Weir, C Counsell, M McDowall, A Gunkel, M Dennis
J.Neurol.Neurosurg.Psychiatry 2003;74: 447-451

Towards a national system for monitoring the quality of hospital based stroke services
N Weir and M Dennis for the Scottish Stroke Outcomes Study Group
Stroke 2001;32:1415-1421

Presentations to learned societies

Stroke outcomes: useful indicators of the quality of stroke care? (platform)
NU Weir on behalf of the Scottish Stroke Outcomes Group
Association of British Neurologists 1999, London

Stroke outcomes: useful indicators of the quality of stroke care? (platform)
NU Weir on behalf of the Scottish Stroke Outcomes Group
British Association of Stroke Physicians 1999, Nottingham

Are stroke outcomes useful indicators of the quality of stroke care? (platform)
NU Weir on behalf of the Scottish Stroke Outcomes Group
National Casemix Conference 1999, Keele University

The Five Hospitals Study (platform)
NU Weir on behalf of the Scottish Stroke Outcomes Group
The Royal College of Physicians of Edinburgh, 1999

Which diagnostic codes are important for large scale and routine stroke studies? (poster)
NU Weir, MA McDowell, AJ Gunkel, MS Dennis
European Stroke Conference 1997, Amsterdam, Holland
Abstract in Cerebrovascular Diseases 1997; 7 (suppl 4): 71 A

Stroke League Tables: are they likely to reflect differences in quality? (poster)
N Weir on behalf of the Scottish Stroke Outcomes Group
European Stroke Conference 1999, Venice, Italy
Abstract in Cerebrovascular Diseases 1999: 9 (suppl 1): 115

References

Aboderin, I. and Venables, G. (1996). Stroke management in Europe. Pan European consensus meeting on stroke management. *Journal of Internal Medicine* 249, 173-180.

Altman, D. G. (1999). Practical statistics for medical research. Chapman & Hall/CRC: London.

Altman, D. G. and Royston, P. (2000). What do we mean by validating a prognostic model? *Stat.Med.* 19, 453-473.

Anderson, C. S., Jamrozik, K. D., Burvill, P. W., Chakera, T. M. H., Johnson, G. A., and Stewart-Wynne, E. G. (1993). Ascertaining the true incidence of stroke: experience from the Perth Community Stroke Study, 1989-90. *Med J Aust* 158, 80-84.

Anderson, P. (1999). Popularising hospital performance data. *BMJ* 318, 1772.

Anonymous. (1989). Stroke. Towards better management. The Royal College of Physicians of London: London.

Anonymous. (1990). Precise clinical summaries: the source of high quality SMR data. Guidelines for clinicians. Summary of recommendations. Edinburgh, Information & Statistics Division, Common Services Agency for the Scottish Health Service.

Anonymous (1993). Dicing with death rates. *Lancet* 341, 1183-1184.

Anonymous (1994). Royal College of Physicians Stroke Audit Package Introductory Booklet. Royal College of Physicians of London: London.

Anonymous (1995). Core Patient Profile Information in Scottish Hospitals (COPPISH) Scottish Morbidity Record Data Manual. ISD, NHS in Scotland: Edinburgh.

Anonymous (1996). Significance tests. In 'An introduction to medical statistics'. (Ed. M. Bland.) pp. 133-151. Oxford University Press: Oxford,UK.

Anonymous (1997). Patterns and trends in male mortality. In 'Health Inequalities'. (F. Drever and M. WhiteheadEds.) pp. 95-107. Office of National Statistics, The Stationary Office: London.

Anonymous (1998a). Fifty years ago. The new NHS: message to the medical profession from the Minister of Health. *BMJ* 317, 56.

Anonymous (1998b). Stroke Care: reducing the burden of disease. The Stroke Association: London.

Anonymous (1999). Comparing groups - continuous data. In 'Practical statistics for medical research'. (Ed. D. G. Altman.) pp. 179-228. Chapman & Hall/CRC: Boca Raton.

Anonymous. (2000). Clinical indicators (league tables) - a discussion paper. p17. London, British Medical Association.

Anonymous. (2001). Coronary Heart Disease/Stroke Task Force Report. The Scottish Executive, Edinburgh

Anonymous. (2002). Coronary Heart Disease and Stroke. Strategy for Scotland. The Scottish Executive, Edinburgh.

Anonymous. (2003a) Scottish Stroke Services Audit: report of an audit on the organisation of services for stroke patients 1997-1998. Glasgow, Royal College of Physicians & Surgeons of Glasgow.

Anonymous. (2003b) Draft clinical standards for stroke services. http://www.clinicalstandards.org/pdf/draftstand/Draft%20stroke.pdf. NHS Quality Improvement Scotland, Edinburgh.

Antiplatelet Trialists' Collaboration. (1988). Secondary prevention of vascular disease by prolonged antiplatelet treatment. *BMJ* 296, 320-321.

Antiplatelet Trialists' Collaboration. (1994). Collaborative overview of randomised trials of antiplatelet therapy - 1: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 308, 81-106.

Antithrombotic Trialists' Collaboration (2002). Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 324, 71-86.

Aron, D. C., Harper, D. L., Shepardson, L. B., and Rosenthal, G. E. (1998). Impact of risk-adjusting cesarean delivery rates when reporting hospital performance. *JAMA* 279, 1968-1972.

Asplund, K., Hulter-Asberg, K., Norrving, B., Stegmayr, B., Terent, A., Wester, P. O., Riks-Stroke Collaboration (2003). Riks-stroke - a Swedish national quality register for stroke care. *Cerebrovasc.Dis.* 15 Suppl 1, 5-7.

Asplund, K., Tuomilehto, J., Stegmayr, B., Wester, P. O., and Tunstall-Pedoe, H. (1988). Diagnostic Criteria and Quality Control of the Registration of Stroke Events in the MONICA Project. *Acta Med Scand* 26-39.

Atrial Fibrillation Investigators. (1994). Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation, analysis of pooled data from five randomised controlled trials. *Archives of Internal Medicine* 154, 1449-1457.

Austin, P. C. and Naylor, C. D. (2001). A comparison of a Bayesian vs. a frequentist method for profiling hospital performance. *Journal of Evaluation in Clinical Practice* 7, 35-45.

Baird, A. E., Dashe, J., Connor, A., Burzynski, C., Schlaug, G., and Warach, S. (2000). Comparison of Retrospective and Prospective Measurements of the National Institutes of Health Stroke Scale. *Cerebrovasc.Dis.* 10, 80-81.

Baker, R. and Fraser, R. C. (1995). Development of review criteria: linking guidelines and assessment of quality. *BMJ* 311, 370-373.

Bamford, J., Sandercock, P., Dennis, M., Burn, J., and Warlow, C. (1990a). A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project 1981-86. 2. Incidence, case fatality rates and overall outcome at one year of cerebral infarction, primary intracerebral and subarachnoid haemorrhage. *J.Neurol.Neurosurg.Psychiatry* 53, 16-22.

Bamford, J., Dennis, M., Sandercock, P., Burn, J., and Warlow, C. (1990b). The frequency, causes and timing of death within 30 days of a first stroke: the Oxfordshire Community Stroke Project. *J.Neurol.Neurosurg.Psychiatry* 53, 824-829.

Bamford, J., Sandercock, P., Dennis, M., Warlow, C., Jones, L., McPherson, K., Vessey, M., Fowler, G., Molyneux, A., and Hughes, T. (1988). A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project 1981-86. 1. Methodology, demography and incident cases of first-ever stroke. *J.Neurol.Neurosurg.Psychiatry* 51, 1373-1380.

Bamford, J. M., Sandercock, P. A., Warlow, C. P., and Slattery, J. (1989). Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 20, 828.

Barclay, S., Todd, C., Finlay, I., Grande, G., and Wyatt, P. (2002). Not another questionnaire! Maximising the response rate, predicting non-response and assessing non-response bias in postal questionnaire studies of GPs. *Family Practice* 19, 105-111.

Barer, D., Ellul, J., and Watkins, C. (1996). Correcting outcome data for casemix in stroke medicine. Structures and processes should be audited, rather than outcomes. *BMJ* 313, 1005-1006.

Barer, D. H. (1989). The natural history and functional consequences of dysphagia after hemispheric stroke. *J.Neurol.Neurosurg.Psychiatry* 52, 236-241.

Barer, D. H. and Mitchell, J. R. (1989). Predicting the outcome of acute stroke: do multivariate models help? *Q.J.Med.* 70, 27-39.

Berwick, D. M. (1996). A primer on leading the improvement of systems. *BMJ* 312, 619-622.

Berwick, D. M., Enthoven, A., and Bunker, J. P. (1992a). Quality management in the NHS: the doctor's role I. *BMJ* 304, 235-239.

Berwick, D. M., Enthoven, A., and Bunker, J. P. (1992b). Quality management in the NHS: the doctor's role II. *BMJ* 304, 304-308.

Berwick, D. M. and Wald, D. L. (1990). Hospital leaders' opinions of the HCFA mortality data. *JAMA* 263, 247-249.

Bilsker, D. and Goldner, E. M. (2002). Routine outcome measurement by mental health-care providers: is it worth doing? *Lancet* 360, 1689-1690.

Bindman, A. B. (1999). Can physician profiles be trusted? *JAMA* 281, 2142-2143.

Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 312, 1215-1218.

Bland, J. M. and Altman, D. G. (1998). Bayesians and frequentists. *BMJ* 317, 1151.

Blumberg, M. S. (1991). Biased estimates of expected acute myocardial infarction mortality using MedisGroups Admission Severity Groups. *JAMA* 265, 2965-2970.

Blumenthal, D. (1994). The variation phenomenon in 1994. *N.Engl.J.Med.* 331, 1017-1018.

Blumenthal, D. (1996a). Quality of Health Care. Part 1: Quality of Care - What is it? *N.Engl.J.Med.* 335, 891-894.

Blumenthal, D. (1996b). Quality of Health Care. Part 4: The origins of the quality of care debate. *N.Engl.J.Med.* 335, 1146-1149.

Bobrow, R. S. (2000). The unintended consequences of measuring quality on the quality of care. *N.Engl.J.Med.* 342, 519.

Bogousslavsky, J., Regli, F., Uske, A., and Maeder, P. (1991). Early spontaneous hematoma in cerebral infarct: is primary cerebral haemorrhage overdiagnosed? *Neurology* 41, 837-840.

Bonita, R. (2002). Epidemiology of stroke. *Lancet* 339, 342-344.

Braitman, L. E. and Davidoff, F. (1996). Predicting clinical states in individual patients. *Annals of Internal Medicine* 125, 406-412.

Brazier, J. E., Walters, S. J., Nicholl, J. P., and Kohler, B. (1996). Using the SF-36 and Euroquol on an elderly population. *Quality of Life Research* 5, 195-204.

Brennan, P. and Silman, A. (1992). Statistical methods for assessing observer variability in clinical measures. *BMJ* 304, 1491-1494.

Brennan, T. A., Leape, L. L., Laird, N. M., Hebert, L., Localio, A. R., Lawthers, A. G., Newhouse, J. P., Weiler, P. C., and Hiatt, H. H. (1991). Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N.Engl.J.Med.* 324, 370-376.

Britton, M., Hindmarsh, T., Murray, V., and Tyden, S. A. (1984). Diagnostic errors discovered by CT in patients with suspected stroke. *Neurology* 34, 1504-1507.

Brook, R. H., McGlynn, E. A., and Cleary, P. D. (1996). Quality in Health Care. Part 2: Measuring quality of care. *N.Engl.J.Med.* 335, 966-970.

Brosseau, L., Potvin, L., Philippe, P., and Boulanger, Y-L. (1996). Post-stroke inpatient rehabilitation. II. Predicting discharge disposition. *American Journal of Physical Medicine & Rehabilitation* 75, 431-436.

Brown, M. M. (2003). Carotid angioplasty and stenting: are they therapeutic alternatives? *Cerebrovasc.Dis.* 11 Supple 1, 112-118.

Burn, J. P. (1992). Reliability of the modified Rankin Scale. *Stroke* 23, 438.

Burns, F. (1998). Information for Health. An information strategy for the modern NHS 1998-2005. A national strategy for local implementation. Leeds, NHS Executive.

Bushnell, C. D., Johnston, D. C. C., and Goldstein, L. B. (2001). Retrospective assessment of initial stroke severity. Comparison of the NIH Stroke Scale and the Canadian Neurological Scale. *Stroke* 32, 656-660.

Campion, F. X. and Rosenblatt, M. S. (1996). Quality assurance and medical outcomes in the era of cost containment. *Surgical Clinics of North America* 76, 139-159.

Candelise, L., Pinardi, G., and Aritzu, E. (1990). Stroke disability assessed by telephone interivew: reliability and reproducibility. *Journal of Neurology* 237, 143.

Caplan, R. A., Posner, K. L., and Cheney, F. W. (1991). Effect of outcome on physician judgements of appropriateness of care. *JAMA* 265, 1957-1960.

Cargill, V., Cohen, D., Kroenke, K., and Neuhauser, D. (1986). Ongoing patient randomization: an innovation in medical care research. *Health Serv.Res.* 21, 663-678.

Chassin, M. R., Hannan, E. L., and DeBuono, B. A. (1996). Benefits and hazards of reporting medical outcomes publicly. *N.Engl.J.Med.* 334, 394-398.

Clarke, A and McKee, M. (1992). The consultant episode: an unhelpful measure. *BMJ* 305, 1307-1308.

Clinical Outcome Working Group. (1993). Clinical Outcome Measures Report. The Scottish Office Clinical Resource and Audit Group, Edinburgh.

Clinical Outcome Working Group. (1994). Clinical Outcome Indicators. The Scottish Office Clinical Resource and Audit Group, Edinburgh.

Clinical Outcome Working Group. (1998). Clinical Outcome Indicators. The Scottish Office Clinical Resource and Audit Group, Edinburgh.

Clinical Outcome Working Group. (1999). Clinical Outcome Indicators. The Scottish Office Clinical Resource and Audit Group, Edinburgh.

Clinical Standards Advisory Group. (1995) Urgent and emergency admissions to hospital. HMSO, London.

Collins, R., Peto, R., MacMahon, S., Hebert, P., Fiebach, N. H., Eberlein, K. A., Godwin, J., Qizilbash, N., Taylor, J. O., and Hennekens, C. H. (1990). Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 335, 827-838.

Counsell, C. and Dennis, M. (2001). Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc.Dis.* 12, 159-170.

Counsell, C., Dennis, M., McDowall, M., and Warlow, C. (2002). Predicting outcome after acute and subacute stroke. Development and validation of new prognostic models. *Stroke* 33, 1041-1047.

Counsell, C. E. (1998) The prediction of outcome in patients with acute stroke. Univeristy of Cambridge,UK. [Thesis]

Crombie, I. K. and Davies, H. T. O. (1998). Beyond health outcomes: the advantages of measuring process. *J.Eval.Clin.Pract.* 4, 31-38.

Czlonkowska, A., Milewska, D., and Ryglewicz, D. (2003). The Polish experience in early stroke care. *Cerebrovasc.Dis.* 15 Suppl 1, 14-15.

D'Olhaberriague, L., Litvan, I., Mitsias, P., and Mansbach, H. H. (1996). A reappraisal of reliability and validity studies in stroke. *Stroke* 27, 2331-2336.

Dans, P. E., Weiner, J. P., and Otter, S. E. (1985). Peer Review Organisations. Promises and potential pitfalls. *N.Engl.J.Med.* 313, 1131-1137.

Davenport, R. J. and Dennis, M. S. (1996). Measuring the process of care is not always straightforward. *BMJ* 312, 185.

Davenport, R. J., Dennis, M. S., and Warlow, C. P. (1995). Improving the recording of the clinical assessment of stroke patients using a clerking proforma. *Age and Aging* 24, 43-48.

Davenport, R. J., Dennis, M. S., Wellwood, I., and Warlow, C. P. (1996a). Complications after acute stroke. *Stroke* 27, 415-420.

Davenport, R. J., Dennis, M. S., and Warlow, C. P. (1996b). Effect of correcting outcome data for case mix: an example from stroke medicine. *BMJ* 312, 1503-1505.

Davenport, R. J., Dennis, M. S., and Warlow, C. P. (1996c). The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. *Health Bulletin* 54 , 402-405.

Davies, H. T. and Shields, A. V. (1999). Public trust and accountability for clinical performance: lessons from the national press reportage of the Bristol hearing. *J.Eval.Clin.Pract.* 5, 335-342.

Davies, H. T. O. and Crombie, I. K. (1997). Interpreting health outcomes. *J.Eval.Clin.Pract.* 3, 187-199.

Davies, H. T. O. and Lampel, J. (1998). Trust in performance indicators? *Quality in Health Care* 7, 159-162.

De Haan, R., Horn, J. S., Limburg, M., Van Der Meulen, J., and Bossuyt, P. (1993). A comparison of five stroke scales with measures of disability, handicap, and quality of life. *Stroke* 24, 1178-1181.

Demlo, L. K. and Campbell, P. M. (1981). Improving hospital discharge data: lessons from the National Hospital Discharge Survey. *Medical Care* 19, 1030-1040.

Dennis, M. (2000). Stroke services: the good, the bad and the ... *Journal of the Royal College of Physicians of London* 34, 92-96.

Dennis, M. (2001). FOOD (Feed Or Ordinary Diet): A family of randomized trials evaluating feeding policies for patients admitted to hospital with recent stroke. *Cerebrovasc. Dis.* 11, 32.

Dennis, M. and Warlow, C. (1991). Strategy for stroke. *BMJ* 636-638.

Dennis, M., Wellwood, I., and Warlow, C. (1997a). Are simple questions a valid measure of outcome after stroke? *Cerebrovasc. Dis.* 7, 22-27.

Dennis, M., Wellwood, I., O'Rourke, S., MacHale, S, and Warlow, C. (1997b). How reliable are simple questions in assessing outcome after stroke? *Cerebrovasc. Dis.* 7, 19-21.

Dennis, M. S., Bamford, J. M., Molyneux, A. J., and Warlow, C. P. (1987). Rapid resolution of signs of primary intracerebral haemorrhage in computed tomograms of the brain. *BMJ (Clin Res Ed)* 295, 379-381.

Department of Health (2000). An organisation with memory: report of an expert group on learning from adverse events in the NHS. The Stationary Office: London.

Department of Health Expert Advisory Group on Cancer. (1995) A policy framework for commissioning cancer services. DOH, London

Dewey, H. M., Donnan, G. A., Freeman, E. J., Sharples, C. M., Macdonell, R. A., McNeil, J. J., and Thrift, A. G. (1999). Interrater reliability of the National Institutes of Health Stroke Scale: rating by neurologists and nurses in a community-based stroke incidence study. *Cerebrovasc.Dis.* 9, 323-327.

Dillman, D. A. (1978). Mail and telephone surveys - the total design method. J. Wiley & Sons: New York.

Doll, H., McPherson, K., Davies, J., Flood, A., Smith, J., Williams, G., Ginzler, M., Petticrew, M., and Black, N. (1991). Reliability of questionnaire responses as compared with interview in the elderly: views of the outcome of transurethral resection of the prostate. *Social Science and Medicine* 33, 1303-1308.

Dombovy, M. L., Basford, J. R., Whisnant, J. P., and Bergstralh, E. J. (1987). Disability and use of rehabilitation services following stroke in Rochester, Minnesota 1975-1979. *Stroke* 18, 830-836.

Donabedian, A. (1988). The quality of care. How can it be assessed? *JAMA* 260, 1743-1748.

Donnan, G. A. (1992). Investigation of patients with stroke and transient ischaemic attack. *Lancet* 339, 879-880.

Dorman, P. J., Slattery, J., Farrell, B., Dennis, M. S., Sandercock, P. A. G., and the United Kingdom collaborators in the International Stroke Trial (1997a). A randomised comparison of the Euroquol and Short Form-36 after stroke. *BMJ* 315, 461.

Dorman, P. J., Waddell, F., Slattery, J., Dennis, M., and Sandercock, P. (1997b). Is the EuroQol a valid measure of health-related quality of life after stroke? *Stroke* 28, 1876-1882.

Dubois, R. W., Rogers, W. H., Moxley, J. H., Draper, D., and Brook, R. H. (1987). Hospital inpatient mortality. Is it a predictor of quality? *N.Engl.J.Med.* 317, 1674-1680.

Duncan, G., Ritchie, L.C., Jamieson, D.M., and McLean, M.A. (1995). Acute stroke in South Ayrshire: comparative study of pre and post stroke units. *Health Bulletin (Edinb).* 53, 159-166

Duncan, P. W., Horner, R. D., Reker, D. M., Samsa, G. P., Hoening, H., Hamilton, B., LaClair, B. J., and Dudley, T. K. (2002a). Adherence to postacute rehabilitation guidelines is associated with functional recovery in stroke. *Stroke* 33, 167-178.

Duncan, P. W., Reker, D. M., Horner, R. D., Samsa, G. P., Hoening, H., LaClair, B. J., and Dudley, T. K. (2002b). Performance of a mail-administered version of a stroke-specific outcome measure, the Stroke Impact Scale. *Clinical Rehabilitation* 16, 493-505.

Duncan, P. W., Lai, S. M., Tyler, D., Perera, S., Reker, D. M., and Studenski, S. (2002c). Evaluation of proxy responses to the Stroke Impact Scale. *Stroke* 33, 2593-2599.

Duncan, P. W., Jorgensen, H. S., and Wade, D. T. (2000). Outcome measures in acute stroke trials. A systematic review and some recommendations to improve practice. *Stroke* 31, 1429-1438.

Duncan, P. W., Lai, S. M., Bode, R. K., Perera, S., and DeRosa, J. (2003). Stroke Impact Scale-16: A brief assessment of physical function. *Neurology* 60, 291-296.

Ebrahim, S. and Redfern, J. (1999) Stroke care - a matter of chance: a national survey of stroke services. The Stroke Association.

Ellekjaer, H., Holmen, J., Kruger, O., and Terent, A. (1999). Identification of incident stroke in Norway. Hospital discharge data compared with a population-based stroke register. *Stroke* 30, 56-60.

Ellwood, P. M. (1988). Shattuck lecture-outcomes management. A technology of patient experience. *N.Engl.J.Med.* 318, 1549-1556.

English, T. A., Bailey, A. R., Dark, J., and Williams, W. G. (1984). The UK Cardiac Surgical Register 1977-82. *BMJ* 289, 1205-1208.

Epstein, A. M. (1998). Rolling down the runway. The challenges ahead for quality report cards. *JAMA* 279, 1691-1696.

European Carotid Surgery Trialists' Collaborative Group. (1991). MRC European Carotid Surgery Trial: interim results for symptomatic patients with severe (70-99%) or with mild (0-29%) carotid stenosis. *Lancet* 337, 1235-1243.

Evans, A., Perez, I., Harraf, F., Melbourn, A., Steadman, J., Donaldson, N., and Kalra, L. (2001). Can differences in management processes explain different outcomes between stroke unit and stroke-team care? *Lancet* 358, 1586-1592.

Ferro, J. M., Falcao, I., Rodrigues, G., Canhao, P., Melo, T. P., Oliveira, V., Pinto, A. N., Crespo, M., and Salgado, A. V. (1996). Diagnosis of transient ischaemic attack by the nonneurologist. *Stroke* 27, 2225-2229.

Ferro, J. M., Pinto, A. N., Falcao, I., Rodrigues, G., Ferreira, J., Falcao, F., Azevedo, E., Canhao, P., Melo, T. P., Rosas, M. J., Oliveira, V., and Salgado, A. V. (1998). Diagnosis of stroke by the nonneurologist. A validation study. *Stroke* 29, 1106-1109.

Fine, L. G., Keogh, B. E., Cretin, S., Orlando, M., Gould, M. M., and for the Nuffield-Rand Cardiac Surgery Demonstration Project Group (2003). How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK Cardiac Surgery Experience. *BMJ* 326, 25-28.

Fleiss, J. L., Cohen, J., and Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psych Bull* 72, 323-327.

Frain, J. P., Frain, A. E., and Carr, P. H. (1996). Experience of medical senior house officers in preparing discharge summaries. *BMJ* 312, 350.

Frater, A. and Costain, D. (1992). Any better? Outcome measures in medical audit. *BMJ* 304, 519-520.

Fullerton, K. J., Mackenzie, G., and Stout, R. W. (1988). Prognostic indices in stroke. *Q.J.Med.* 66, 147-162.

Gautam, P., Macduff, C., Brown, I., and Squair, J. (1996). Unplanned readmissions of elderly patients. *Health Bulletin* 54, 449-457.

Gillum, L. A. and Johnston, S. C. (2001). Characteristics of academic medical centers and ischemic stroke outcomes. *Stroke* 32, 2137-2142.

Giuffrida, A., Gravelle, H., and Roland, M. (1999). Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ* 319, 94-98.

Gjorup, T., Kelbaek, H., Stenbygard, L., and Sorensen, F. (1986). Effect of knowledge of serum enzyme concentrations on doctors' interpretation of electrocardiographic manifestations in suspected acute myocardial infarction. *BMJ* 292, 27.

Glader, E. L., Stegmayr, B., Johansson, L., Hulter-Asberg, K., and Wester, P. O. (2001). Differences in long-term outcome between patients treated in stroke units and in general wards: a 2-year follow-up of stroke patients in Sweden. *Stroke* 32, 2124-2130.

Gladman, J. R., Harwood, D. M., and Barer, D. H. (1992). Predicting the outcome of acute stroke: prospective evaluation of five multivariate models and comparison with simple methods. *J.Neurol.Neurosurg.Psychiatry* 55, 347-351.

Glance, L. G., Osler, T., and Shinozaki, T. (2000). Effect of varying the casemix on the standardised mortality ratio and W statistic. A simulation study. *Chest* 117, 1112-1117.

Goldstein, H. and Myers, K. (1996). Freedom of information: towards a code of ethics for performance indicators. *Research Intelligence* 57, 12-16.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of instiutional performance. *Journal of the Royal Statistical Society* 159, 385-443.

Goldstein, L. B. (1998). Accuracy of ICD-9-CM coding for the identification of patients with acute ischaemic stroke. Effect of modifier codes. *Stroke* 29, 1602-1604.

Goldstein, L. B. and Chilukuri, V. (1997). Retrospective assessment of initial stroke severity with the Canadian Neurological Scale. *Stroke* 28, 1181-1184.

Gompertz, P., Dennis, M., Hopkins, A., and Ebrahim, S. (1994). Development and reliability of the Royal College of Physicians Stroke Audit Form. *Age and Aging* 23, 378-383.

Gompertz, P., Pound, P., Briffa, J., and Ebrahim, S. (1995). How useful are non-random comparisons of outcomes and quality of care in purchasing hospital stroke services? *Age and Aging* 24, 137-141.

Gompertz, P. H., Irwin, P., Morris, R., Lowe, D., Rutledge, Z., Rudd, A. G., and Pearson, M. G. (2001). Reliability and validity of the Intercollegiate Stroke Audit Package. *J.Eval.Clin.Pract.* 7, 1-11.

Green, J., Passman, L. J., and Wintfeld, N. (1991). Analyzing hospital mortality. The consequences of diversity in patient mix. *JAMA* 265, 1849-1853.

Green, J. and Wintfeld, N. (1993). How accurate are hospital discharge data for evaluating effectiveness of care? *Medical Care* 31, 719-731.

Green, J. and Wintfeld, N. (1995). Report cards on cardiac surgeons. Assessing New York State's approach. *N.Engl.J.Med.* 332, 1229-1232.

Green, J., Wintfeld, N., Krasner, M., and Wells, C. (1997). In search of America's best hospitals. The promise and reality of quality assessment. *JAMA* 277, 1152-1155.

Green, J., Wintfeld, N., Sharkey, P., and Passman, L. J. (1990). The importance of severity of illness in assessing hospital mortality. *JAMA* 263, 241-246.

Green, S. B. and Byar, D. P. (1984). Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat.Med.* 3, 361-370.

Greenfield, S., Cretin, S., Worthman, L. G., Dorey, F. J., Solomon, N. E., and Goldberg, G. A. (1981). Comparison of a criteria map to a criteria list in quality-of-care assessment for patients with chest pain: the relation of each to outcome. *Medical Care* 19, 255-272.

Grimshaw, J. and Russell, I. (1993). Achieving health gain through clinical guidelines. 1:developing scientifically valid guidelines. *Quality in Health Care* 2, 243-248.

Gulliford, M. C., Petruckevitch, A., and Burney, P. G. J. (1991). Hospital case notes and medical audit: evaluation of non-response. *BMJ* 302, 1128-1129.

Hancock, R. J., Oddy, M., Saweirs, W. M., and Court, B. (1997). The RCP stroke audit package in practice. *J.R.Coll.Physicians Lond* 31, 74-78.

Hannan, E. L., Racz, M. J., Jollis, J. G., and Peterson, E. D. (1997). Using medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv.Res.* 31, 659-678.

Harley, K. and Jones, C. (1996). Quality of Scottish Morbidity Record (SMR) data. *Health Bulletin* 54, 410-417.

Harrell, F. E., Jr., Lee, K. L., and Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat.Med.* 15, 361-387.

Hartz.A.J. and Kuhn, E. M. (1994). Comparing hospitals that perfrom coronary artery bypass surgery: the effect of outcome measures and data sources. *Am.J.Public Health* 84, 1609-1614.

Hartz.A.J., Kuhn, E. M., and Krakauer, H. (1997). The relationship of the value of outcome comparisons to the number of patients per provider. *International Journal for Quality in Health Care* 9, 247-254.

Hasan, M., Meara, R. J., and Bhowmick, B. K. (1995). The quality of diagnostic coding in cerebrovascular disease. *International Journal for Quality in Health Care* 7, 407-410.

Heart Protection Study Collaborative Group (2002). MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 360, 7-22.

Hennekens, C. H. and Buring, J. E. (1987). Cohort Stuides. In 'Epidemiology in Medicine'. pp. 153-177. (Little, Brown and Company: Boston/Toronto.)

Henon, H., Godefroy, O., Leys, D., Mounier-Vehier, F., Lucas, C., Rondepierre, P., Duhamel, A., and Pruvo, J. P. (1995). Early predictors of death and disability after acute cerebral ishemic event. *Stroke* 26, 392-398.

Heuschmann, P. U., Berger, K., Misselwitz, B., Hermanek, P., Leffmann, C., Adelmann, M., Buecker-Nott, H-J., Rother, J., Neundoerfer, B., Kolominsky-Rabas, P. L., for the German Stroke Registers Study Group and for the Competence Net Stroke (2003). Frequency of thrombolytic therapy in patients with acute ischemic stroke and the risk of in-hospital mortality. The German Stroke Registers Study Group. *Stroke* 34, 1106-1113.

Hoening, H., Duncan, P. W., Horner, R. D., Reker, D. M., Samsa, G. P., Dudley, T. K., and Hamilton, B. B. (2002). Structure, process and outcomes in stroke rehabilitation. *Medical Care* 40, 1036-1047.

Hoening, H., Sloane, R., Horner, R. D., Zolkewitz, M., and Reker, D. (2001). Differences in rehabilitation services and outcomes among stroke patients cared for in veterans hospitals. *Health Serv.Res.* 35, 1293-1318.

Hoeymans, N., Feskens, E. J., Van Den Bos, G. A., and Kromhout, D. (1998). Non-response bias in a study of cardiovascular diseases, functional status and self-rated health among elderly men. *Age Ageing* 27, 35-40.

Hofer, T. P. and Hayward, R. A. (1996). Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Medical Care* 34, 737-753.

Hollis, S., Yates, D. W., Woodford, M., and Foster, P. (1995). Standardised comparison of performance indicators in trauma: a new approach to case-mix correction. *The Journal of Trauma: Injury, Infection and Critical Care* 38, 763-766.

Holloway, R. G., Vickrey, B. G., Benesch, C., Hinchey, J. A., and Bieber, J. (2001). Development of performance measures for acute ischaemic stroke. *Stroke* 32, 2058-2074.

Hopkins, A. (1996). Clinical audit: time for a reappraisal? *Journal of the Royal College of Physicians of London* 30, 415-425.

Hosmer, D. W. and Lemeshow, S. (1995). Confidence interval estimates of an index of quality performance based on logistic regression models. *Stat.Med.* 14, 2161-2172.

Hsia, D. C., Krushat, W. M., Fagan, A. B., Tebbutt, J. A., and Kusserow, R. P. (1988). Accuracy of diagnostic coding for Medicare patients under the Prospective-Payment System. *N.Engl.J.Med.* 318, 352-355.

Hulka, B. S. and Romm, F. J. (1979). Peer review in ambulatory care: use of explicit criteria and implicit judgements. *Medical Care* 17, 1-73.

Iezzoni, L. (1990). Using administrative diagnostic data to assess the quality of hospital care. The pitfalls and potential uses of ICD-9-CM. *International Journal of Technology Assessment in Health Care* 6, 272-281.

Iezzoni, L. (1997a). Risk adjustment for measuring health care outcomes. (Health Administration Press: Chicago, Illinois.)

Iezzoni, L., Daley, J., Heeren, T., Foley, S. M., Hughes, J. S., Fisher, E. S., Duncan, C. C., and Coffman, G. A. (1994). Using administrative data to screen hospitals for high complication rates. *Inquiry* 31, 40-55.

Iezzoni, L. I. (1994). Using risk-adjusted outcomes to assess clinical practice: an overview of issues pertaining to risk adjustment. *Annals of Thoracic Surgery* 58, 1822-1826.

Iezzoni, L. I. (1996). 100 apples divided by 15 red herrings: cautionary tale from the mid-19th century on comparing hospital mortality rates. *Annals of Internal Medicine* 124, 1079-1085.

Iezzoni, L. I. (1997b). The risks of risk adjustment. *JAMA* 278, 1600-1607.

Iezzoni, L. I., Foley, S. M., Daley, J., Hughes, J., Fisher, E. S., and Heeren, T. (1992). Comorbidities, complicatoins and coding bias. *JAMA* 267, 2197-2203.

Indredavik, B, Bakke, F., Slordahl, S. A., Rokseth, R., and Haheim, L. L. (1999). Treatment in a combined acute and rehabilitation stroke unit. Which aspects are most important? *Stroke* 30, 917-923.

International Stroke Trial Collaborative Group. (1997). The International Stroke Trial (IST); a randomised trial of aspirin, subcutaneous heparin, both or neither among 19435 patients with acute ischaemic stroke. *Lancet* 349, 1569-1581.

Irwin, P. and Rudd, A. (1998). Casemix and process indicators of outcome in stroke. The Royal College of Physicians minimum data set for stroke. *J.R.Coll.Physicians Lond* 32, 442-444.

Jakovljevic, D, Sarti, C., Sivenius, J., Torppa, J., Mahonen, M., Immonen-Raiha, P., Kaarsalo, E., Alhainen, K., Kuulasmaa, K., Tuomilehto, J., Puska, P., and Salomaa, V. (2001). Socioeconomic status and ischaemic stroke. *Stroke* 32, 1492-1498.

Jencks, S. F., Cuerdon, T., Burwen, D. R., Fleming, B., Houck, P. M., Kussmaul, A. E., Nilasena, D. S., Ordin, D. L., and Arday, D. R. (2000). Quality of medical care delivered to medicare beneficiaries. A profile at state and national levels. *JAMA* 284, 1670-1676.

Jencks, S. F., Williams, D. K., and Kay, T. L. (1988a). Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. *JAMA* 260, 2240-2246.

Jencks, S. F., Daley, J., Draper, D., Thomas, N., Lenhart, G., and Walker, J. (1988b). Interpreting hospital mortality data. The role of clinical risk adjustment. *JAMA* 260, 3611-3616.

Jencks, S. F., Huff, E. D., and Cuerdon, T. (2003). Changes in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. *JAMA* 289, 305-312.

Jessee, W. F. and Schranz, C. M. (1990). Medicare mortality rates and hospital quality: are they related? *Qual.Assur.Health Care* 2, 137-144.

Jollis, J. G. and Romano, P. S. (1998). Pennsylvania's focus on heart attack - grading the scorecard. *JAMA* 338, 983-987.

Jorgensen, H. S., Nakayama, H., Raaschou, H. O., Vive-Larsen, J., Stoier, M., and Olsen, T. S. (1995). Outcome and time course of recovery in stroke. Part II: time course of recovery. The Copenhagen Study Group. *Arch.Phys.Med.Rehabil.* 76, 406-411.

Justice, A. C., Covinsky, K. E., and Berlin, J. A. (1999). Assessing the generalisability of prognostic information. *Annals of Internal Medicine* 130, 515-524.

Kahn, K. L., Rogers, W. H., Rubenstein, L. V., Sherwood, M. J., Reinisch, E. J., Keeler, E. B., Draper, D., Kosecoff, J., and Brook, R. H. (1990). Measuring quality

of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. *JAMA* 264, 1969-1973.

Kasner, S. E., Chalela, J. A., Luciano, J. M., Cucchiara, B. L., Raps, E. C., McGarvey, M. L., Conroy, M. B., and Localio, A. R. (1999). Reliability and validity of estimating the NIH Stroke Scale score from medical records. *Stroke* 30, 1534-1537.

Kay, R., Wong, K. S., Perez, G., and Woo, J. (1997). Dichotomizing stroke outcomes based on self-reported dependency. *Neurology* 49, 1694-1696.

Kendrick, S. (2001). Using all the evidence: towards a truly intelligent National Health Service. *Health Bull.(Edinb.)* 59, 71-76.

Kendrick, S. and Clarke, J. (1993). The Scottish record linkage system. *Health Bulletin* 51, 72-79.

King's Fund Consensus Conference. (1988). Treatment of stroke. *BMJ* 297, 126-128.

Knaus, W. A., Draper, E. A., Wagner, D. P., and Zimmerman, J. E. (1986). An evaluation of outcome from intensive care in major medical centers. *Annals of Internal Medicine* 104 , 410-418.

Kohli, H. S. and Knill-Jones, R. P. (1992). How accurate are SMR1 (Scottish Morbidity Record 1) data? *Health Bull.(Edinb.)* 50, 14-23.

Kothari, R. U., Brott, T., Broderick, J. P., and Hamilton, C. A. (1995). Emergency physicians. Accuracy in the diagnosis of stroke. *Stroke* 26, 2238-2241.

Kramer, A. M., Kowalsky, J. C., Lin, M., Grigsby, J., Hughes, R., and Steiner, J. F. (2000). Outcome and utilisation differences for older persons with stroke in HMO and fee-for-service systems. *Journal of the American Geriatric Society* 48, 726-734.

Kunst, A. E., del Rios, M., Groenhof, F., Mackenbach, J. P., and for the European Union Working Group on Socioeconmic Inequalities in Health (1998). Socioeconomic inequalities in stroke mortality among middle-aged men. An international overview. *Stroke* 29, 2285-2291.

LaClair, B. J., Reker, D. M., Duncan, P. W., Horner, R. D., and Hoening, H. (2001). Stroke care:a method for measuring compliance with AHCPR guidelines. *American Journal of Physical Medicine & Rehabilitation* 80, 235-242.

Laffel, G. and Blumenthal, D. (1989). The case for using industrial quality management science in health care organisations. *JAMA* 262, 2869-2873.

Langford, I. H. (1997). Bayesian analysis should be used instead of league tables of performance. *BMJ* 314, 73-74.

Langhorne, P. (1995). Developing comprehensive stroke services: an evidence-based approach. *Postgraduate Medical Journal* 71, 733-737.

Langhorne, P. and Dennis, M. (1998). Stroke units: an evidence based approach. BMJ Books: London.

Langhorne, P., Li Pak Tong, B., and Stott, D. J. (1999). Association between physiological homeostasis and early recovery after stroke. *Stroke* 30, 2526-2527.

Langhorne, P. and Pollock, A (2002). What are the components of effective stroke unit care? *Age Ageing* 31, 365-371.

Langhorne, P., Wagenaar, R., and Partridge, C. (1996). Physiotherapy after stroke: more is better? *Physiotherapy Research International* 1, 75-88.

Langhorne, P., Williams, B. O., Gilchrist, W., and Howie, K. (1993). Do stroke units save lives? *Lancet* 342, 395-398.

Laupacis, A., Sekar, N., and Stiell, I. G. (1997). Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 277, 488-494.

Leatherman, S. and McCarthy, D. (1999). Public disclosure of health care performance reports: experience, evidence and issues for policy. *International Journal for Quality in Health Care* 11, 93-105.

Leibson, C. L., Naessens, J. M., Brown, R. D., and Whisnant, J. P. (1994). Accuracy of hospital discharge abstracts for identifying stroke. *Stroke* 25, 2348-2355.

Lembcke, P. A. (1967). Medical auditing by scientific methods illustrated by major female pelvic surgery. *JAMA* 162, 646-655.

Libman, R. B., Wirkowski, E., Alvir, J., and Rao, T. H. (1995). Conditions that mimic stroke in the emergency department. Implications for acute stroke trials. *Ach.Neurol.* 52, 1119-1122.

Lindley, R. I., Amayo, E. O., Marshall, J., Sandercock, P. A., Dennis, M., and Warlow, C. P. (1995). Hospital services for patients with acute stroke in the United Kingdom: the Stroke Association survey of consultant opinion. *Age Ageing* 24, 525-532.

Lindley, R. I., Waddell, F., Livingstone, M., Sandercock, P., Dennis, M. S., Slattery, J., Smith, B., and Warlow, C. (1994). Can simple questions assess outcome after stroke? *Cerebrovasc. Dis.* 4, 314-324.

Liu, L., Reeder, B., Shuaib, A., and Mazagri, R. (1999). Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. *Cerebrovasc.Dis.* 9, 224-230.

Long, A. F. (1997b). Measuring health outcomes - requirements for routine practice. In 'Health outcomes and evaluations: context, concepts and successful applications'. (A. F. Long and E. Bitzer Eds.) pp. 47-57. Universtiy of Leeds: Leeds.

Long, A. F. (1997a). The role of health outcomes in health care evaluation. In 'Health outcomes and evaluation: context, concepts and successful applications'. (A. F. Long and E. Bitzer Eds.) pp. 33-43. University of Leeds: Leeds.

Luft, H. S. and Hunt, S. S. (1986). Evaluating individual hospital quality through outcome statistics. *JAMA* 255, 2780-2784.

Lyons, T. F. and Payne, B. C. (1974). The relationship of physicians' medical recording performance to their medical care performance. *Medical Care* 12, 714-720.

Mallinson, S. (1998). The Short-Form 36 and older people: some problems encountered when using postal administration. *J.Epidemiol.Community Health* 52, 324-328.

Malmgren, R., Bamford, J., Warlow, C., Sandercock, P., and Slattery, J. (1989). Projecting the number of patients with first ever stroke and patients newly handicapped by stroke in England and Wales. *BMJ* 298, 656-660.

Mannion, R. and Davies, H. T. O. (2002). Reporting health care performance: learning from the past, prospects for the future. *J.Eval.Clin.Pract.* 8, 215-228.

Mannion, R. and Goddard, M. (2003). Public disclosure of comparative clinical performance data: lessons from the Scottish experience. *J.Eval.Clin.Pract.* 9, 277-286.

Mant, J. (2001). Process versus outcome indicators in the assessment of quality of health care. *International Journal for Quality in Health Care* 13, 475-480.

Mant, J. and Hicks, N. (1995). Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 311, 793-796.

Mant, J. and Hicks, N. R. (1996). Assessing quality of care: what are the implications of the potential lack of sensitivity of outcome measures to differences in quality? *J.Eval.Clin.Pract.* 2, 243-248.

Mant, J., Mant, F., and Winner, S. (1997). How good is routine information? Validation of coding for acute stroke in Oxford hospitals. *Health Trends* 29, 96-99.

Marshall, E. C. and Spiegelhalter, D. J. (1998). Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ* 316, 1701-1705.

Marshall, M. N., Shekelle, P. G., Leatherman, S., and Brook, R. H. (2000). The public release of performance data. What do we expect to gain? A review of the evidence. *JAMA* 283, 1866-1874.

Martin, J., Meltzer, H., and Elliot, D. (1988). OPCS surveys of disability in Great Britain. The prevalence of disability amongst adults. HMSO: London.

Maxwell, R. J. (1984). Quality assessment in health. *BMJ* 288, 1470-1472.

Maxwell, R. J. (1992). Dimensions of quality revisited: from thought to action. *Qual.Health Care* 1 , 171-177.

McAuliffe, W. E. (1978). Studies of Process-Outcome correlations in medical care evaluations: a critique. *Medical Care* 16, 907-930.

McCloone, P. (1994) Carstairs scores for Scottish postcode sectors from the 1991 Census. Public Health Research Unit, University of Glasgow, Glasgow.

McDowell, I. and Newell, C. (1996). Measuring Health. A guide to rating scales and questionnaires.

McKee, M. and Hunter, D. (1995). Mortality league tables: do they inform or mislead? *Quality in Health Care* 4, 5-12.

McKee, M. and James, P. (1997). Using routine data to evaluate quality of care in British hospitals. *Medical Care* 35, OS102-OS111.

McKee, M., Rafferty, A-M., and Aiken, L. (1997). Measuring hospital performance: are we asking the right questions? *Journal of the Royal Society of Medicine* 90, 187-191.

McKenna, K., Tooth, L., Strong, J., Ottenbacher, K., Connell, J., and Clearly, M. (2002). Predicting discharge outcomes for stroke patients in Australia. *American Journal of Physical Medicine & Rehabilitation* 81, 47-56.

McNaughton, H. (1996). Stroke audit in a New Zealand hospital. *N.Z.Med.J.* 109, 257-260.

McNaughton, H., McPherson, K., Taylor, W., and Weatherall, M. (2003). Relationship between process and outcome in stroke care. *Stroke* 34, 713-717.

McPherson, K., Wennberg, J. E., Hovind, O. B., and Clifford, P. (1982). Small area variations in the use of common surgical procedures: an international comparison of New England, England and Norway. *N.Engl.J.Med.* 307, 1310-1314.

Miles, A., Bentley, P., Grey, J., and Polychronis, A. (1995). Purchasing quality in clinical practice; what on Earth do we mean? *J.Eval.Clin.Pract.* 1, 87-95.

Milne, R. and Clarke, A (1990). Can readmission rates be used as an outcome indicator? *BMJ* 301, 1139-1140.

Mitchell, J. B., Ballard, D. J., Whisnant, J. P., Ammering, C. J., Samsa, G. P., and Matchar, D. B. (1996). What role do neurologists play in determining the costs and outcomes of stroke patients? *Stroke* 27, 1937-1943.

Modan, B. and Wagener, D. K. (1992). Some epidemiological aspects of stroke: mortality/morbidity trends, age, sex, race, socioeconomic status. *Stroke* 23, 1230-1236.

Mohammed, M. A., Cheng, K. K., Rouse, A., and Marshall, T. (2001). Bristol, Shipman, and clinical governance: Shewart's forgotten lessons. *Lancet* 357, 463-467.

Moreno, R. and Apolone, G. (1997). Impact of different customisation strategies in the performance of a general severity score. *Critical Care Medicine* 25, 2001-2008.

Murray, C. J. L. and Lopez, A. D. (1997a). Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 349, 1269-1276.

Murray, C. J. L. and Lopez, A. D. (1997b). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet* 349, 1436-1442.

Naylor, C. D. (1995). Grey zones of clinical practice:some limits to evidence-based medicine. *Lancet* 345, 840-842.

Neuhauser, D. (1990). Ernest Amory Codman,M.D., and end results of medical care. *International Journal of Technology Assessment in Health Care* 6, 307-325.

News at Ten and Independent Television News. 9-12-1994. London, UK.

NHS Executive. (1999) Quality and performance in the NHS:clinical indicators. BMA Books, London.

Normand, S-L. T., Glickman, M. E., Sharma, R. G. V. R. K., and McNeil, B. J. (1996). Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients. Results from the Cooperative Cardiovascular Project. *JAMA* 275 , 1322-1328.

Norris, J. W. and Hachinski, V. C. (1982). Misdiagnosis of stroke. *Lancet* 1, 328-331.

North American Symptomatic Carotid Endarterectomy Trial Collaborators. (1991). Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N.Engl.J.Med.* 325, 445-453.

Noseworthy, J. H., Ebers, G. C., Vandervoort, M. K., Farquhar, R. E., Yetisir, E., and Roberts, R. (1994). The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 44, 16-20.

O'Connor, G. T., Quinton, H. B., Traven, N. D., Ramunno, L. D., Dodds, T. A., Marciniak, T. A., and Wennberg, J. E. (1999). Geographic variation in the treatment of acute myocardial infarction: the Cooperative Cardiovascular Project. *JAMA* 17, 627-633.

O'Mahoney, P. G., Rodgers, H., Thomson, R. G., Dobson, R., and James, O. F. W. (1998). Is the SF-36 suitable for assessing health status of older stroke patients? *Age Aging* 27, 19-22.

Omoigui, N. A., Miller, D. P., Brown, K. J., Annan, K., Cosgrove, D 3rd., Lytle, P., Loop, F., and Topol, E. J. (1996). Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation* 93, 27-33.

Orchard, C. (1994). Comparing health outcomes. *BMJ* 308, 1493-1496.

Palmer, R. H. (1997). Process-based measures of quality: the need for detailed clinical data in large health care databases. *Annals of Internal Medicine* 127, 733-738.

Panayiotou, B. N., Fotherby, M. D., Potter, J. F., and Castleden, C. M. (1993). The accuracy fo diagnostic coding of cerebrovascular disease. *Medical Audit News* 3, 153-155.

Parker, C. J., Dewey, M. E., and on behalf of the TOTAL Study Group (2000). Assessing research outcomes by postal questionnaire with telephone follow-up. *Int.J.Epidemiol.* 29, 1065-1069.

Parry, G. J., Gould, C. R., McCabe, C. J., and Tarnow-Mordi, W. O. (1998). Annual league tables of mortality in neonatal intensive care units: longitudinal study. International Neonatal Network and the Scottish Neonatal Consultants and Nurses Collaborative Study Group. *BMJ* 316, 1931-1935.

Patel, A. R., Gray, G., Lang, G. D., Baillie, F. G. H., Fleming, L., and Wilson, G. M. (1976a). Scottish Hospital Morbidity Data. 1. Errors in diagnostic returns. *Health Bull.(Edinb.)* 34, 215-220.

Patel, A. R., Hutcheon, A. W., and Wilson, G. M. (1976b). Scottish Hospital Morbidity Data. 3. Some suggestions for improvement of the accuracy. *Health Bull.(Edinb.)* 34, 227-228.

Peterson, E. D., DeLong, E. R., Jollis, J. G., Muhlbaier, L. H., and Mark, D. B. (1998). The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. *Journal of the American College of Cardiology* 32, 993-999.

Picavet, H. S. (2001). National health surveys by mail or home interview: effects on response. *J.Epidemiol.Community Health* 55, 408-413.

Pound, P., Sabin, C., and Ebrahim, S. (1999). Observing the process of care: a stroke unit, elderly care unit and general medical ward compared. *Age Ageing* 28, 433-440.

PROGRESS Collaborative Group (2001). Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. *Lancet* 358, 1033-1041.

Radical Statistics Health Group (1995). NHS "indicators of success": what do they tell us? *BMJ* 310 , 1045-1050.

Relman, A. S. (1988). Assessment and accountability. The third revolution in medical care. *N.Engl.J.Med.* 319, 1220-1222.

Retchin, S. M., Brown, R. S., Yeh, S. C., Chu, D., and Moreno, L. (1997). Outcomes of stroke patients in Medicare fee for service and managed care. *JAMA* 278, 119-124.

Ricci, S., Celani, M. G., La Rosa, F., Vitali, R., Duca, E., Ferraguzzi, R., Paolotti, M., Seppoloni, D., Caputo, N., and Chiurulla, C. (1991). SEPIVAC: a community-based study of stroke incidence in Umbria, Italy. *J.Neurol.Neurosurg.Psychiatry* 54, 695-698.

Roberts, L. and Counsell, C. (1998). Assessment of clinical outcomes in acute stroke trials. *Stroke* 29, 986-991.

Rockall, T. A., Logan, R. F. A., Devlin, H. B., Northfield, T. C., for the National Audit of Acute Upper Gastrointestinal Haemorrhage (1995). Variation in outcome after acute upper gastrointestinal haemorrhage. *Lancet* 346, 346-350.

Rodgers, H., Dennis, M., Cohen, D., Rudd, A., on behalf of the British Association of Stroke Physicians (2003). British Association of Stroke Physicians: benchmarking survey of stroke services. *Age Ageing* 32, 211-217.

Romano, P. S. and Mark, D. H. (1994). Bias in the coding of hospital discharge data and its implications for quality assessment. *Medical Care* 32, 81-90.

Rosenthal, G. E., Shah, A., Way, L. E., and Harper, D. L. (1998). Variations in standardized hospital mortality rates for six common medical diagnoses. Implications for profiling hospital quality. *Medical Care* 36, 955-964.

Rothwell, P. M. (2000). Interpretation of variations in outcome in audit of clinical interventions. *Lancet* 355, 4-5.

Rothwell, P. M., Eliasziw, M., Gutnikov, S. A., Fox, A. J., Taylor, D. W., Mayberg, M. R., Warlow, C. P., Barnett, H. J., for the Carotid Trialists' Collaboration (2003). Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet* 361, 107-116.

Rothwell, PM. and Warlow, C. (1995). Is self-audit reliable? *Lancet* 346, 1623.

Rowley, G. and Fielding, K. (1991). Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet* 337, 535-538.

Rudd, A. G., Irwin, P., Rutledge, Z., Lowe, D., Wade, D., Morris, R., and Pearson, M. G. (1999). The national sentinel audit for stroke: a tool for raising standards of care. *J.R.Coll.Physicians Lond* 33, 460-464.

Rudd, A. G., Irwin, P., Rutledge, Z., Lowe, D., Wade, D. T., and Pearson, M. (2001a). Regional variations in stroke care in England, Wales and Northern Ireland: results from the National Sentinel Audit of Stroke. *Clinical Rehabilitation* 15, 562-572.

Rudd, A. G., Lowe, D., Irwin, P., Rutledge, Z., Pearson, M., on behalf of the Intercollegiate Stroke Working Party (2001b). National stroke audit: a tool for change? *Quality in Health Care* 10, 141-151.

Rudd, A. G. and Pearson, M. (2002). National stroke audit. *Clinical Medicine* 2, 496-498.

Ryan, M. P. and Dodd, W. (1993). The measurement of process and the role of information technology. *Proceedings of the Royal Society of Edinburgh* 101B, 31-53.

Sackett, D. L., Haynes, R. B., and Tugwell, P. (1985). Clinical Epidemiology A basic science for clinical medicine. Little, Brown and Company: Boston/Toronto.

Schmulling, S., Grond, M., and Rudolf, J. (1998). Training as a prerequisite for reliable use of NIH Stroke Scale. *Stroke* 29, 1258-1259.

Schwartz, J. S. and Lurie, N. (1990). Assessment of medical outcomes. New opportunities for achieving a long sought-after objective. *International Journal of Technology Assessment in Health Care* 6, 333-339.

Scott, C. (1961). Research on mail surveys. *Journal of the Royal Statistical Society* 124, 143-195.

Scottish Intercollegiate Guidelines Network (SIGN). (1997a). Management of patients with stroke I. Assessment, investigation, immediate management and secondary prevention. SIGN: Edinburgh.

Scottish Intercollegiate Guidelines Network (SIGN). (1997b). Management of patients with stroke: II. Management of carotid stenosis and carotid endarterectomy. SIGN: Edinburgh.

Scottish Intercollegiate Guidelines Network (SIGN) (1997c). Management of patients with stroke: III. Identification and management of dysphagia. SIGN: Edinburgh.

Scottish Intercollegiate Guidelines Network (SIGN) (1998). Management of patients with stroke: IV. Rehabilitation, prevention and management of complications, and discharge planning. SIGN: Edinburgh.

Secretary of State for Health, Wales Northern Ireland and Scotland. (1989) Working for Patients. HMSO, London.

Secretary of State for Scotland. (1997) Designed to Care. The Stationery Office, Edinburgh.

Shapiro, M. F, Park, R. E., Keesey, J., and Brook, R. H. (1994). The effect of alternative case-mix adjustments on mortality differences between municipal and voluntary hospitals in New York City. *Health Serv.Res.* 29, 95-112.

Sheikh, K. and Mattingly, S. (1981). Investigating non-response bias in mail surveys. *Journal of Epidemiology and Community Health* 35, 293-296.

Shinar, D., Gross, C. R., Mohr, J. P., Caplan, L. R., Price, T. R., Wolf, P. A., Hier, D. B., Kase, C. S., Fishman, I. G., and Wolf, C. L. (1985). Interobserver variability in the assessment of neurologic history and examination in the Stroke Data Bank. *Arch.Neurol.* 42, 557-565.

Shu, B-P., Lemeshow, S., Hosmer, D. W., Klar, J., Avrunin, J., and Teres, D. (1996). Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customisation: a simulation study. *Critical Care Medicine* 24, 57-63.

Siemiatycki, J. (1979). A comparison of mail, telephone, and home interview strategies for household health surveys. *Am.J.Public Health* 69, 238-245.

Signorini, D. F. and Weir, N. U. (1999). Any variability in outcome comparisons adjusted for casemix must be accounted for. *BMJ* 318, 128.

Silber, J. H. and Rosenbaum, P. R. (1997). A spurious correlation between hospital mortality and complication rates. *Medical Care* 35, OS77-OS92.

Silber, J. H., Rosenbaum, P. R., Schwartz, J. S., Ross, R. N., and Williams, S. V. (1995). Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. *JAMA* 274, 317-323.

Skilbeck, C. E., Wade, D. T., Hewer, R. L., and Wood, V. A. (1983). Recovery after stroke. *J.Neurol.Neurosurg.Psychiatry* 46, 5-8.

Sloan, M., Kreiger, N., and James, B. (1997). Improving response rates among doctors: randomised trial. *BMJ* 315, 1136.

Smeeth, L., Fletcher, A. E., Stirling, S., Nunes, M., Breeze, E., Ng, E., Bulpitt, C. J., and Jones, D. (2001). Randomised comparison of three methods of administering a screening questionnaire to elderly people: findings from the MRC trial of the assessment and management of older people in the community. *BMJ* 323, 1-7.

Smith, E. J., Ward, A. J., and Smith, D. (1990). Trauma scoring methods. *Br.J.Hosp.Med.* 44, 114-118.

Smith, P. (1990). The use of performance indicators in the public sector. *Journal of the Royal Statistical Society* 153, 53-72.

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 18, 277-310.

Smith, R. (1998). All changed, changed utterly. British medicine will be transformed by the Bristol case. *BMJ* 316, 1917-1918.

Smith, T. (1994). Clinical outcome measures: medical emergency readmissions. *Health Bulletin* 52, 333-341.

Smith, W. C., Crombie, I. K., Campion, P. D., and Knox, J. D. (1985). Comparison of response rates to a postal questionnaire from a general practice and a research unit. *BMJ(Clin.Res.Ed)* 291, 1483-1485.

Smithard, D. G., O'Neill, P. A., Park, C., Morris, J., Wyatt, R., England, R., and Martin, D. F. (1996). Complications and outcome after acute stroke. Does dysphagia matter? *Stroke* 27, 1200-1204.

Spiegelhalter, D. J. (1999). Surgical audit: statistical lessons from Nightingale and Codman. *Journal of the Royal Statistical Society* 162, 45-58.

Sprague, L. (2002). Contracting for quality; Medicare's quality improvement organisations. *NHPF Issue Brief* 1-15.

Starfield, B., Steinwachs, D., Morris, I., Bause, G., Siebert, B. A., and Westin, C. (1979). Concordance between medical records and observations regarding information on coordination of care. *Medical Care* 17, 758-766.

Stegmayr, B. and Asplund, K. (1992). Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. *Neuroepidemiology* 11 , 204-213.

Stegmayr, B., Asplund, K., Hulter-Asberg, K., Norrving, B., Peltonen, M., Terent, A., and Wester, P. O. (1999). Stroke units in their natural habitat: can the results of randomised trials be reproduced in routine clinical practice? *Stroke* 30, 709-714.

Steiner, M. M., Brainin, M., the participants in the Austrian Stroke Registry for Acute Stroke Units (2003). The quality of acute stroke units on a nation-wide level: the Austrian Stroke Registry for acute stroke units. *European Journal of Neurology* 10, 353-360.

Sterne, J. A. C. and Davey Smith, G. (2001). Sifting the evidence - what's wrong with significance tests? *BMJ* 322, 226-231.

Stingele, R., Bluhmki, E., Hacke, W. (2001) Bootstrap statistics of ECASS 2 data: just another post hoc analysis of a negative stroke trial? *Cerebrovasc. Dis.* 11, 30-33.

Streiner, D. L. and Norman, G. R. (1989). Health measurement scales. A practical guide to their development and use. Oxford University Press: Oxford.

Stroke Unit Trialists' Collaboration. (1997a). Collaborative systematic review of the randomised trials of organised (stroke unit) care after stroke. *BMJ* 314, 1151-1158.

Stroke Unit Trialists' Collaboration. (1997b). How do stroke units improve patient outcomes? A collaborative systematic review of the randomised trials. *Stroke* 28, 2139-2144.

Taub, N. A., Wolfe, C. D., Richardson, E., and Burney, P. G. (1994). Predicting the disability of first-time stroke sufferers at 1 year. 12-month follow-up of a population-based cohort in southeast England. *Stroke* 25, 352-357.

The Heart Outcome Prevention Evaluation Study Investigators (2000). Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high risk patients. *N.Engl.J.Med.* 342, 145-153.

The Intercollegiate Working Party for Stroke (2000). National Clinical Guidelines for Stroke. Royal College of Physicians of London: London.

The Intercollegiate Working Party for Stroke (2002). National Clinical Guidelines for Stroke: a concise update. *Clinical Medicine* 2, 231-233.

The National Institute for Neurological Disorders and Stroke rt-PA Stroke Study Group (1995). Tissue plasminogen activator for acute ischaemic stroke. *N.Engl.J.Med.* 333, 1581-1587.

The Secretary of State for Health. (1997) The new NHS modern and dependable. DOH, London.

Thomas, J. W. and Hofer, T. P. (1998). Research evidence on the validity of risk-adjusted mortality rate as a measure of hospital quality of care. *Medical Care Research and Review* 55, 371-404.

Thomas, J. W. and Hofer, T. P. (1999). Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. *Medical Care* 37, 83-92.

Thomas, J. W. and Holloway, J. J. (1991). Investigating early readmission as an indicator for quality of care studies. *Medical Care* 29, 377-394.

Thomas, J. W., Holloway, J. J., and Guire, K. E. (1993). Validating risk-adjusted mortality as an indicator for quality of care. *Inquiry* 30, 6-22.

Thomson, R. G., McElroy, H., and Kazandjian, V. A. (1997). Maryland Hospital Quality Indicator Project in the United Kingdom: an approach for promoting continuous quality improvement. *Quality in Health Care* 6, 1-7.

Thorvaldsen, P., Asplund, K., Kuulasmaa, K., Rajakangas, A., Schroll, M., for the WHO MONICA Project (1995). Stroke incidence, case fatality, and mortality in the WHO MONICA Project. *Stroke* 26, 361-367.

Treurniet, H. F., Essink-Bot, M-L., Mackenbach, J. P., and van der Maas, P. J. (1997). Health-related quality of life: an indicator of quality of care? *Quality of Life Research* 6 , 363-369.

Tymms, P. and Wiggins, A. (2000). Schools' experience of league tables should make doctors think again. *BMJ* 321, 1467.

van den Bos, G. A. M., Smits, J. P. J. M., Westert, G. P., and van Straten, A. (2002). Socioeconomic variations in the course of stroke: unequal health outcomes, equal care? *Journal of Epidemiology and Community Health* 56, 943-948.

van Gijn, J. (1992). Measurement of outcome in stroke prevention trials. *Cerebrovasc. Dis.* 2, 23-34.

van Straten, A., van der Meulen, J. H. P., van Crevel, H., Habbema, J. D. F., and Limburg, M. (1997). Quality of hospital care for stroke patients in the Netherlands. *Cerebrovasc. Dis.* 7, 251-257.

von Arbin, M., Britton, M., de Faire, U., Helmers, C., Miah, K., and Murray, V. (1981). Accuracy of bedside diagnosis in stroke. *Stroke* 12, 288-293.

Wade, D. T. (1995a). Personal physical disability. In 'Measurement in neurological rehabilitation'. (Ed. Anonymous.) pp. 70-82. Oxford University Press: Oxford.

Wade, D. T. (1995b). Activities of daily living (ADL) and extended ADL tests. In 'Measurement in neurological rehabilitation'. (Ed. Anonymous.) pp. 175-194. Oxford University Press: Oxford.

Wade, D. T. and Hewer, R. L. (1985). Outlook after an acute stroke: urinary incontinence and loss of consciousness compared in 532 patients. *Q.J.Med.* 56, 601-608.

Wade, D. T. and Hewer, R. L. (1987). Functional abilities after stroke: measurement, natural history and prognosis. *J.Neurol.Neurosurg.Psychiatry* 50, 177-182.

Walshe, K. and Higgins, J. (2002). The use and impact of inquiries in the NHS. *BMJ* 325, 895-900.

Wardlaw, J. M. (1994). Controversies in management: computed tomography provides accurate diagnosis. *BMJ* 309, 1498-1499.

Wardlaw, J. M., del Zoppo, G., and Yamaguchi, T. (2003). Thrombolysis for acute ischaemic stroke (Cochrane Review). In 'The Cochrane Library'. Update Software: Oxford.

Warlow, C. P., Dennis, M. S., Hankey, G. J., Sandercock, P. A. G., Bamford, J. M., and Wardlaw, J. (1996). Reducing the burden of stroke and improving public health. In 'Stroke: a practical guide to management'. pp. 632-649. Blackwell Science: Oxford,UK.

Wasson, J. H., Sox, H. C., Neff, R. K., and Goldman, L. (1985). Clinical prediction rules. Applications and methodological standards. *N.Engl.J.Med.* 313, 793-798.

Wattigney, W. A., Croft, J. B., Mensah, G. A., Alberts, M. J., Shephard, T. J., Gorelick, P. B., Nilasena, D. S., Hess, D. C., Walker, M. D., Hanley, D. F., Shwayder, P., Girgus, M., Neff, L. J., Williams, J. E., LaBarthe, D. R., and Collins, J.

L. (2003). Establishing data elements for the Paul Coverdell National Acute Stroke Registry. *Stroke* 34, 151-156.

Ween, J. E., Alexander, M. P., D'Esposito, M., and Roberts, M. (1996). Incontinence after stroke in a rehabilitation setting: outcome associations and predictive factors. *Neurology* 47, 659-663.

Weir, N. U., Sandercock, P. A. G., Lewis, S. C., Signorini, D. F., Warlow, C. P., on behalf of the IST Collaborative Group (2001). Variations between countries in outcome after stroke in the International Stroke Trial (IST). *Stroke* 32, 1370-1377.

Wen, S. W., Hernandez, R., and Naylor, C. D. (1995). Pitfalls in nonrandomized outcomes studies. The case of incidental appendectomy with open cholecystectomy. *JAMA* 274, 1687-1691.

Wennberg, J. E. (2002). Unwarranted variations in healthcare delivery: implications for academic medical centres. *BMJ* 325, 961-964.

Wennberg, J. E., Freeman, J. L., and Culp, W. J. (1987). Are hospital services rationed in New Haven or over-utilised in Boston? *Lancet* 1, 1185-1189.

Westgren, M., Divon, M., Greenspoon, J., and Paul, R. (1986). Missing hospital records: a confounding variable in retrospective studies. *American Journal of Obstetrics and Gynaecology* 155, 269-271.

Wilkin, A. and McColl, I. (1987). Surgical audit: the clinician's view. *Theoretical Surgery* 1, 195-206.

Wilkin, D. (1987). Conceptual problems in dependency research. *Social Science and Medicine* 24, 867-873.

Williams, J. G. and Mann, R. Y. (2002). Hospital episode statistics: time for clinicians to get involved? *Clinical Medicine* 2, 34-37.

Williams, L. S., Yilmaz, E. Y., and Lopez-Yunez, A. M. (2000). Retrospective assessment of initial stroke severity with the NIH stroke scale. *Stroke* 31, 858-862.

Wilson, K. and Roe, B. (1998). Interviewing older people by telephone following initial contact by postal survey. *J.Adv.Nurs.* 27, 575-581.

Wolfe, C. D., Tilling, K., Beech, R., Rudd, A. G., for the European BIOMED Study of Stroke Care Group (1999). Variations in case fatality and dependency from stroke in western and central Europe. *Stroke* 30, 350-356.

Wolfe, C. D. A. (2000). The impact of stroke. *British Medical Bulletin* 56, 275-286.

World Health Organisation (1980). The International Classification of Impairments, Disabilities and Handicaps. World Health Organisation: Geneva.

World Health Organisation (2001). The International Classification of Function, Disability and Health. World Health Organisation: Geneva.

Wyatt, J. C. (1995). Hospital information management: the need for clinical leadership. *BMJ* 311, 175-180.

Wyatt, J. C. and Altman, D. G. (1995). Commentary: Prognostic models: clinically useful or quickly forgotten? *BMJ* 311, 1539-1541.

Yammarino, F. J., Skinner, S. J., and Childers, T. L. (1991). Understanding mail survey response behaviour. *Public Opinion Quarterly* 55, 613-639.

Yao, P., Wiggs, B. R., Gregor, C., Sigurnjak, R., and Dodek, P. (1999). Discordance between physicians and coders in assignment of diagnoses. *International Journal for Quality in Health Care* 11, 147-153.

Yates, D. W., Woodford, M., and Hollis, S. (1992). Preliminary analysis of the care of injured patients in 33 British hospitals: first report of the United Kingdom major trauma outcome study. *BMJ* 305, 737-740.

Yates, J. M. and Davidge, M. G. (1984). Can you measure performance? *BMJ* 288, 1935-1936.

Yeoh, C. and Davies, H. (1993). Clinical coding: completeness and accuracy when doctors take it on. *BMJ* 306, 972.

Younge, P. A., Coats, T. J., Gurney, D., and Kirk, C. J. C. (1997). Interpretation of the Ws statistic: application to an integrated trauma system. *The Journal of Trauma: Injury, Infection and Critical Care* 43, 511-515.

Zalkind, D. L. and Eastaugh, S. R. (1997). Mortality rates as an indicator of hospital quality. *Hosp.Health Serv.Adm* 42, 3-15.

Zuckerman, A. E., Starfield, B., Hochreiter, C., and Kovasznay, B. (1975). Validating the content of pediatric outpatient medical records by means of tape-recording doctor-patient encounters. *Paediatrics* 56, 407-411.